1. Introduction to Fluids
    1. Introduction to Fluid Statics
    2. What Is a Fluid?
    3. Density
    4. Pressure
    5. Variation of Pressure with Depth in a Fluid
    6. Pascal's Principle
    7. Gauge Pressure, Absolute Pressure, and Pressure Measurement
    8. Archimedes' Principle
    9. Cohesion and Adhesion in Liquids: Surface Tension and Capillary Action
    10. Pressures in the Body
2. Fluid Dynamics in the Circulatory System and Cellular Processes
    1. Structure and Function of Blood Vessels
    2. An Overview of Blood
    3. Blood Flow, Blood Pressure, and Resistance
    4. Introduction to Fluid Dynamics and Its Biological and Medical Applications
    5. Flow Rate and Its Relation to Velocity
    6. Viscosity and Laminar Flow; Poiseuille's Law
    7. Bernoulli's Equation
    8. The Most General Applications of Bernoulli's Equation
    9. The Onset of Turbulence
    10. Motion of an Object in a Viscous Fluid
    11. Molecular Transport Phenomena: Diffusion, Osmosis, and Related Processes
    12. The Cell Membrane
3. Electricity in Cells, Muscles and Neurons
    1. Electric Forces in Biology

2. OU Human Physiology: The Action Potential
3. Human Biology Chapter 8.3: Cardiac Muscle and Electrical Activity
4. Nerve Conduction–Electrocardiograms
5. How Neurons Communicate
6. Energy and Metabolism
7. Energy in Living Systems
8. Bis2A 06.0 Energy in Living Systems v1.2
9. Free Energy
10. Potential, Kinetic, Free, and Activation Energy

4. Electric charge and field
    1. Introduction to Electric Charge and Electric Field
    2. Static Electricity and Charge: Conservation of Charge
    3. Conductors and Insulators
    4. Coulomb’s Law
    5. Electric Field: Concept of a Field Revisited
    6. Electric Field Lines: Multiple Charges
    7. Conductors and Electric Fields in Static Equilibrium
    8. Applications of Electrostatics
    9. Introduction to Electric Current, Resistance, and Ohm's Law
5. Electric Potential and Energy
    1. Introduction to Electric Potential and Electric Energy
    2. Electric Potential Energy: Potential Difference
    3. Electric Potential in a Uniform Electric Field
    4. Electrical Potential Due to a Point Charge
    5. Equipotential Lines
    6. Capacitors and Dielectrics
    7. DC Circuits Containing Resistors and Capacitors
    8. Capacitors in Series and Parallel
    9. Energy Stored in Capacitors
6. Electric Circuits

1. Introduction to Circuits and DC Instruments
2. DC Voltmeters and Ammeters
3. Electromotive Force: Terminal Voltage
4. Current
5. Ohm’s Law: Resistance and Simple Circuits
6. Resistance and Resistivity
7. Resistors in Series and Parallel
8. Electric Power and Energy
9. Alternating Current versus Direct Current
10. Electric Hazards and the Human Body
11. DC Circuits Containing Resistors and Capacitors

Introduction to Fluid Statics

class="introduction"

The fluid essential to all life has a beauty of its own. It also helps support the weight of this swimmer. (credit: 12019, Pixabay)

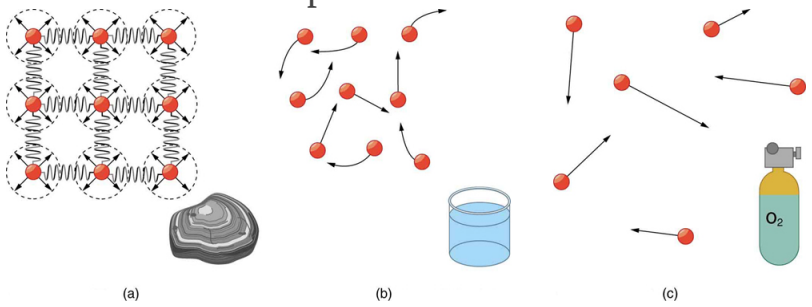


Much of what we value in life is fluid: a breath of fresh winter air; the hot blue flame in our gas cooker; the water we drink, swim in, and bathe in; the blood in our veins. What exactly is a fluid? Can we understand fluids with the laws already presented, or will new laws emerge from their study? The physical characteristics of static or stationary fluids and some of the laws that govern their behavior are the topics of this chapter. Fluid Dynamics and Its Biological and Medical Applications explores aspects of fluid flow.

What Is a Fluid?

- State the common phases of matter.
- Explain the physical characteristics of solids, liquids, and gases.
- Describe the arrangement of atoms in solids, liquids, and gases.

Matter most commonly exists as a solid, liquid, or gas; these states are known as the three common *phases of matter*. Solids have a definite shape and a specific volume, liquids have a definite volume but their shape changes depending on the container in which they are held, and gases have neither a definite shape nor a specific volume as their molecules move to fill the container in which they are held. (See [link].) Liquids and gases are considered to be fluids because they yield to shearing forces, whereas solids resist them. Note that the extent to which fluids yield to shearing forces (and hence flow easily and quickly) depends on a quantity called the viscosity which is discussed in detail in Viscosity and Laminar Flow; Poiseuille's Law. We can understand the phases of matter and what constitutes a fluid by considering the forces between atoms that make up matter in the three phases.



(a) Atoms in a solid always have the same neighbors, held near home by forces represented here by springs. These atoms are essentially in contact with one another. A rock is an example of a solid. This rock retains its shape because of the forces holding its atoms together. (b) Atoms in a liquid are also in close contact but can slide over one another. Forces between them strongly resist attempts to push them closer together and also hold them in close contact.

Water is an example of a liquid. Water can flow, but it also remains in an open container because of the forces between its atoms. (c) Atoms in a gas are separated by distances that are considerably larger than the size of the atoms themselves, and they move about freely. A gas must be held in a closed container to prevent it from moving out freely.

Atoms in *solids* are in close contact, with forces between them that allow the atoms to vibrate but not to change positions with neighboring atoms. (These forces can be thought of as springs that can be stretched or compressed, but not easily broken.) Thus a solid *resists* all types of stress. A solid cannot be easily deformed because the atoms that make up the solid are not able to move about freely. Solids also resist compression, because their atoms form part of a lattice structure in which the atoms are a relatively fixed distance apart. Under compression, the atoms would be forced into one another. Most of the examples we have studied so far have involved solid objects which deform very little when stressed.

Note:
Connections: Submicroscopic Explanation of Solids and Liquids
Atomic and molecular characteristics explain and underlie the macroscopic characteristics of solids and fluids. This submicroscopic explanation is one theme of this text and is highlighted in the Things Great and Small features in Conservation of Momentum. See, for example, microscopic description of collisions and momentum or microscopic description of pressure in a gas. This present section is devoted entirely to the submicroscopic explanation of solids and liquids.

In contrast, *liquids* deform easily when stressed and do not spring back to their original shape once the force is removed because the atoms are free to slide about and change neighbors—that is, they *flow* (so they are a type of fluid), with the molecules held together by their mutual attraction. When a liquid is placed in a container with no lid on, it remains in the container (providing the container has no holes below the surface of the liquid!). Because the atoms are closely packed, liquids, like solids, resist compression.

Atoms in *gases* are separated by distances that are large compared with the size of the atoms. The forces between gas atoms are therefore very weak, except when the atoms collide with one another. Gases thus not only flow (and are therefore considered to be fluids) but they are relatively easy to compress because there is much space and little force between atoms. When placed in an open container gases, unlike liquids, will escape. The major distinction is that gases are easily compressed, whereas liquids are not. We shall generally refer to both gases and liquids simply as **fluids**, and make a distinction between them only when they behave differently.

Note:
PhET Explorations: States of Matter—Basics
Heat, cool, and compress atoms and molecules and watch as they change between solid, liquid, and gas phases.
https://phet.colorado.edu/sims/html/states-of-matter-basics/latest/states-of-matter-basics_en.html

Section Summary

- A fluid is a state of matter that yields to sideways or shearing forces. Liquids and gases are both fluids. Fluid statics is the physics of stationary fluids.

Conceptual Questions

Exercise:

Problem:

What physical characteristic distinguishes a fluid from a solid?

Exercise:

Problem:

Which of the following substances are fluids at room temperature: air, mercury, water, glass?

Exercise:

Problem: Why are gases easier to compress than liquids and solids?

Exercise:

Problem: How do gases differ from liquids?

Glossary

fluids
: liquids and gases; a fluid is a state of matter that yields to shearing forces

Density

- Define density.
- Calculate the mass of a reservoir from its density.
- Compare and contrast the densities of various substances.

Which weighs more, a ton of feathers or a ton of bricks? This old riddle plays with the distinction between mass and density. A ton is a ton, of course; but bricks have much greater density than feathers, and so we are tempted to think of them as heavier. (See [link].)

Density, as you will see, is an important characteristic of substances. It is crucial, for example, in determining whether an object sinks or floats in a fluid. Density is the mass per unit volume of a substance or object. In equation form, density is defined as
Equation:

$$\rho = \frac{m}{V},$$

where the Greek letter ρ (rho) is the symbol for density, m is the mass, and V is the volume occupied by the substance.

Note:
Density
Density is mass per unit volume.
Equation:

$$\rho = \frac{m}{V},$$

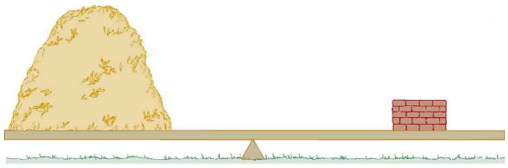
where ρ is the symbol for density, m is the mass, and V is the volume occupied by the substance.

In the riddle regarding the feathers and bricks, the masses are the same, but the volume occupied by the feathers is much greater, since their density is much lower. The SI unit of density is kg/m^3, representative values are given in [link]. The metric system was originally devised so that water would have a density of $1\ \text{g/cm}^3$, equivalent to $10^3\ \text{kg/m}^3$. Thus the basic mass unit, the kilogram, was first devised to be the mass of 1000 mL of water, which has a volume of 1000 cm^3.

Substance	$\rho(10^3\ \text{kg/m}^3 \text{ or } \text{g/mL})$	**Substance**	$\rho(10^3\ \text{kg/m}^3 \text{ or } \text{g/mL})$	**Substance**	$\rho($
Solids		**Liquids**		**Gases**	
Aluminum	2.7	Water (4ºC)	1.000	Air	

Substance	$\rho(10^3\ \mathrm{kg/m^3}$ or $\mathrm{g/mL})$	Substance	$\rho(10^3\ \mathrm{kg/m^3}$ or $\mathrm{g/mL})$	Substance	$\rho($
Brass	8.44	Blood	1.05	Carbon dioxide	
Copper (average)	8.8	Sea water	1.025	Carbon monoxide	
Gold	19.32	Mercury	13.6	Hydrogen	
Iron or steel	7.8	Ethyl alcohol	0.79	Helium	
Lead	11.3	Petrol	0.68	Methane	
Polystyrene	0.10	Glycerin	1.26	Nitrogen	
Tungsten	19.30	Olive oil	0.92	Nitrous oxide	
Uranium	18.70			Oxygen	
Concrete	2.30–3.0			Steam $(100^{\circ}\ \mathrm{C})$	
Cork	0.24				
Glass, common (average)	2.6				
Granite	2.7				
Earth’s crust	3.3				
Wood	0.3–0.9				
Ice (0°C)	0.917				
Bone	1.7–2.0				

Densities of Various Substances



A ton of feathers and a ton of bricks have the same mass, but the feathers make a much bigger pile because they have a much lower density.

As you can see by examining [link], the density of an object may help identify its composition. The density of gold, for example, is about 2.5 times the density of iron, which is about 2.5 times the density of aluminum. Density also reveals something about the phase of the matter and its substructure. Notice that the densities of liquids and solids are roughly comparable, consistent with the fact that their atoms are in close contact. The densities of gases are much less than those of liquids and solids, because the atoms in gases are separated by large amounts of empty space.

Note:
Take-Home Experiment Sugar and Salt
A pile of sugar and a pile of salt look pretty similar, but which weighs more? If the volumes of both piles are the same, any difference in mass is due to their different densities (including the air space between crystals). Which do you think has the greater density? What values did you find? What method did you use to determine these values?

Example:
Calculating the Mass of a Reservoir From Its Volume
A reservoir has a surface area of $50.0\ \text{km}^2$ and an average depth of 40.0 m. What mass of water is held behind the dam? (See [link] for a view of a large reservoir—the Three Gorges Dam site on the Yangtze River in central China.)
Strategy
We can calculate the volume V of the reservoir from its dimensions, and find the density of water ρ in [link]. Then the mass m can be found from the definition of density
Equation:

$$\rho = \frac{m}{V}.$$

Solution
Solving equation $\rho = m/V$ for m gives $m = \rho V$.
The volume V of the reservoir is its surface area A times its average depth h:
Equation:

$$\begin{aligned} V &= \text{Ah} = \left(50.0\ \text{km}^2\right)(40.0\ \text{m}) \\ &= \left[\left(50.0\ \text{km}^2\right)\left(\tfrac{10^3\ \text{m}}{1\ \text{km}}\right)^2\right](40.0\ \text{m}) = 2.00 \times 10^9\ \text{m}^3 \end{aligned}$$

The density of water ρ from [link] is $1.000 \times 10^3\ \text{kg/m}^3$. Substituting V and ρ into the expression for mass gives
Equation:

$$\begin{aligned} m &= \left(1.00 \times 10^3 \text{ kg/m}^3\right)\left(2.00 \times 10^9 \text{ m}^3\right) \\ &= 2.00 \times 10^{12} \text{ kg}. \end{aligned}$$

Discussion
A large reservoir contains a very large mass of water. In this example, the weight of the water in the reservoir is $mg = 1.96 \times 10^{13}$ N, where g is the acceleration due to the Earth's gravity (about 9.80 m/s^2). It is reasonable to ask whether the dam must supply a force equal to this tremendous weight. The answer is no. As we shall see in the following sections, the force the dam must supply can be much smaller than the weight of the water it holds back.



Three Gorges Dam in central China. When completed in 2008, this became the world's largest hydroelectric plant, generating power equivalent to that generated by 22 average-sized nuclear power plants. The concrete dam is 181 m high and 2.3 km across. The reservoir made by this dam is 660 km long. Over 1 million people were displaced by the creation of the reservoir. (credit: Le Grand Portage)

Section Summary

- Density is the mass per unit volume of a substance or object. In equation form, density is defined as
  Equation:

$$\rho = \frac{m}{V}.$$

- The SI unit of density is kg/m^3.

Conceptual Questions

Exercise:

Problem: Approximately how does the density of air vary with altitude?

Exercise:

Problem:

Give an example in which density is used to identify the substance composing an object. Would information in addition to average density be needed to identify the substances in an object composed of more than one material?

Exercise:

Problem:

[link] shows a glass of ice water filled to the brim. Will the water overflow when the ice melts? Explain your answer.



Problems & Exercises

Exercise:

Problem: Gold is sold by the troy ounce (31.103 g). What is the volume of 1 troy ounce of pure gold?

Solution:

$1.610\ \text{cm}^3$

Exercise:

Problem:

Mercury is commonly supplied in flasks containing 34.5 kg (about 76 lb). What is the volume in liters of this much mercury?

Exercise:

Problem:

(a) What is the mass of a deep breath of air having a volume of 2.00 L? (b) Discuss the effect taking such a breath has on your body’s volume and density.

Solution:

(a) 2.58 g

(b) The volume of your body increases by the volume of air you inhale. The average density of your body decreases when you take a deep breath, because the density of air is substantially smaller than the average density of the body before you took the deep breath.

Exercise:

Problem:

A straightforward method of finding the density of an object is to measure its mass and then measure its volume by submerging it in a graduated cylinder. What is the density of a 240-g rock that displaces 89.0 cm^3 of water? (Note that the accuracy and practical applications of this technique are more limited than a variety of others that are based on Archimedes' principle.)

Solution:

$2.70\ \text{g/cm}^3$

Exercise:

Problem:

Suppose you have a coffee mug with a circular cross section and vertical sides (uniform radius). What is its inside radius if it holds 375 g of coffee when filled to a depth of 7.50 cm? Assume coffee has the same density as water.

Exercise:

Problem:

(a) A rectangular gasoline tank can hold 50.0 kg of gasoline when full. What is the depth of the tank if it is 0.500-m wide by 0.900-m long? (b) Discuss whether this gas tank has a reasonable volume for a passenger car.

Solution:

(a) 0.163 m

(b) Equivalent to 19.4 gallons, which is reasonable

Exercise:

Problem:

A trash compactor can reduce the volume of its contents to 0.350 their original value. Neglecting the mass of air expelled, by what factor is the density of the rubbish increased?

Exercise:

Problem:

A 2.50-kg steel gasoline can holds 20.0 L of gasoline when full. What is the average density of the full gas can, taking into account the volume occupied by steel as well as by gasoline?

Solution:

$7.9 \times 10^2\ \text{kg/m}^3$

Exercise:

Problem:

What is the density of 18.0-karat gold that is a mixture of 18 parts gold, 5 parts silver, and 1 part copper? (These values are parts by mass, not volume.) Assume that this is a simple mixture having an average density equal to the weighted densities of its constituents.

Solution:

$15.6\ \text{g/cm}^3$

Exercise:

Problem:

There is relatively little empty space between atoms in solids and liquids, so that the average density of an atom is about the same as matter on a macroscopic scale—approximately $10^3\ \mathrm{kg/m^3}$. The nucleus of an atom has a radius about 10^{-5} that of the atom and contains nearly all the mass of the entire atom. (a) What is the approximate density of a nucleus? (b) One remnant of a supernova, called a neutron star, can have the density of a nucleus. What would be the radius of a neutron star with a mass 10 times that of our Sun (the radius of the Sun is $7 \times 10^8\ \mathrm{m}$)?

Solution:

(a) $10^{18}\ \mathrm{kg/m^3}$

(b) $2 \times 10^4\ \mathrm{m}$

Glossary

density
: the mass per unit volume of a substance or object

Pressure

- Define pressure.
- Explain the relationship between pressure and force.
- Calculate force given pressure and area.

You have no doubt heard the word **pressure** being used in relation to blood (high or low blood pressure) and in relation to the weather (high- and low-pressure weather systems). These are only two of many examples of pressures in fluids. Pressure P is defined as

Equation:

$$P = \frac{F}{A}$$

where F is a force applied to an area A that is perpendicular to the force.

Note:
Pressure
Pressure is defined as the force divided by the area perpendicular to the force over which the force is applied, or

Equation:

$$P = \frac{F}{A}.$$

A given force can have a significantly different effect depending on the area over which the force is exerted, as shown in [link]. The SI unit for pressure is the *pascal*, where

Equation:

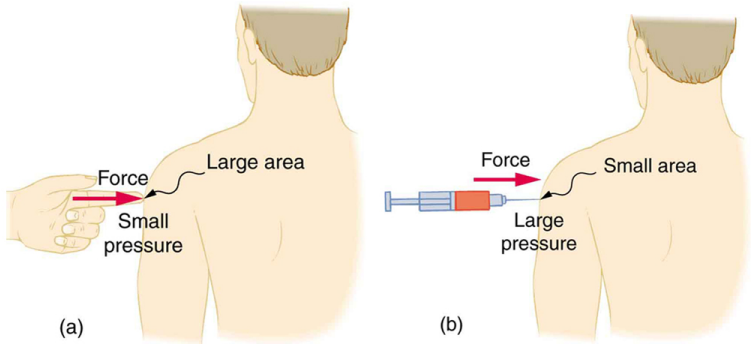
$$1\ \text{Pa} = 1\ \text{N/m}^2.$$

In addition to the pascal, there are many other units for pressure that are in common use. In meteorology, atmospheric pressure is often described in units of millibar (mb), where

Equation:

$$100\ \text{mb} = 1 \times 10^4\ \text{Pa}\ .$$

Pounds per square inch $\left(\text{lb/in}^2 \text{ or psi}\right)$ is still sometimes used as a measure of tire pressure, and millimeters of mercury (mm Hg) is still often used in the measurement of blood pressure. Pressure is defined for all states of matter but is particularly important when discussing fluids.



(a) While the person being poked with the finger might be irritated, the force has little lasting effect. (b) In contrast, the same force applied to an area the size of the sharp end of a needle is great enough to break the skin.

Example:
Calculating Force Exerted by the Air: What Force Does a Pressure Exert?

An astronaut is working outside the International Space Station where the atmospheric pressure is essentially zero. The pressure gauge on her air tank

reads 6.90×10^6 Pa. What force does the air inside the tank exert on the flat end of the cylindrical tank, a disk 0.150 m in diameter?

Strategy

We can find the force exerted from the definition of pressure given in $P = \frac{F}{A}$, provided we can find the area A acted upon.

Solution

By rearranging the definition of pressure to solve for force, we see that

Equation:

$$F = \mathrm{PA}.$$

Here, the pressure P is given, as is the area of the end of the cylinder A, given by $A = \pi r^2$. Thus,

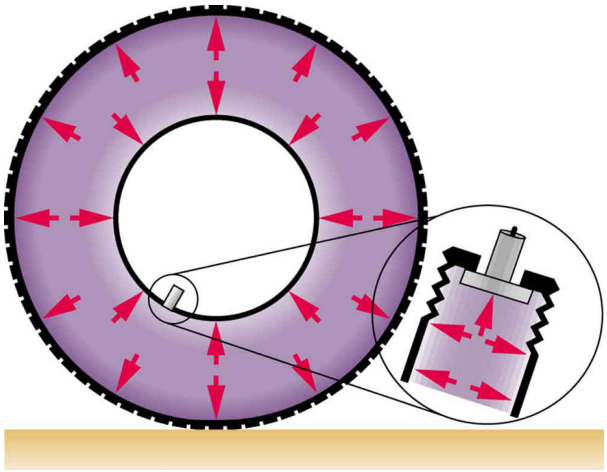
Equation:

$$\begin{aligned} F &= \left(6.90 \times 10^6 \text{ N/m}^2\right)(3.14)(0.0750 \text{ m})^2 \\ &= 1.22 \times 10^5 \text{ N}. \end{aligned}$$

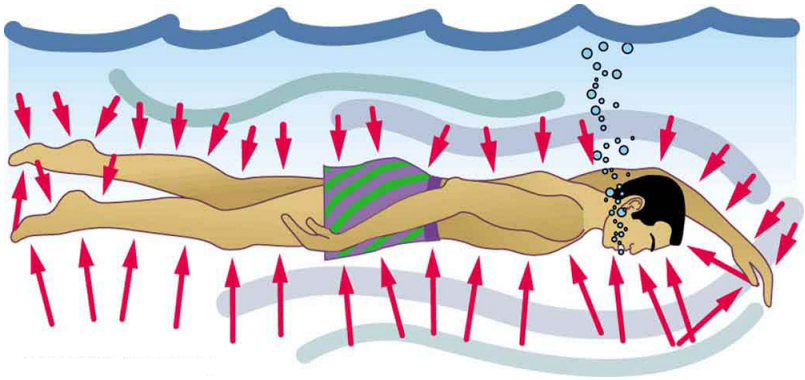
Discussion

Wow! No wonder the tank must be strong. Since we found $F = \mathrm{PA}$, we see that the force exerted by a pressure is directly proportional to the area acted upon as well as the pressure itself.

The force exerted on the end of the tank is perpendicular to its inside surface. This direction is because the force is exerted by a static or stationary fluid. We have already seen that fluids cannot *withstand* shearing (sideways) forces; they cannot *exert* shearing forces, either. Fluid pressure has no direction, being a scalar quantity. The forces due to pressure have well-defined directions: they are always exerted perpendicular to any surface. (See the tire in [link], for example.) Finally, note that pressure is exerted on all surfaces. Swimmers, as well as the tire, feel pressure on all sides. (See [link].)



Pressure inside this tire exerts forces perpendicular to all surfaces it contacts. The arrows give representative directions and magnitudes of the forces exerted at various points. Note that static fluids do not exert shearing forces.



Pressure is exerted on all sides of this swimmer, since the water would flow into the space he occupies if he were not there. The arrows represent the directions and magnitudes of the forces exerted at various points on the swimmer. Note that the forces are larger underneath, due to greater depth,

giving a net upward or buoyant force that is balanced by the weight of the swimmer.

Note:
PhET Explorations: Gas Properties
Pump gas molecules to a box and see what happens as you change the volume, add or remove heat, change gravity, and more. Measure the temperature and pressure, and discover how the properties of the gas vary in relation to each other.

Gas Properties

Section Summary

- Pressure is the force per unit perpendicular area over which the force is applied. In equation form, pressure is defined as
  Equation:

$$P = \frac{F}{A}.$$

- The SI unit of pressure is pascal and $1\ \text{Pa} = 1\ \text{N/m}^2$.

Conceptual Questions

Exercise:

Problem:

How is pressure related to the sharpness of a knife and its ability to cut?

Exercise:

Problem:

Why does a dull hypodermic needle hurt more than a sharp one?

Exercise:

Problem:

The outward force on one end of an air tank was calculated in [link]. How is this force balanced? (The tank does not accelerate, so the force must be balanced.)

Exercise:

Problem:

Why is force exerted by static fluids always perpendicular to a surface?

Exercise:

Problem:

In a remote location near the North Pole, an iceberg floats in a lake. Next to the lake (assume it is not frozen) sits a comparably sized glacier sitting on land. If both chunks of ice should melt due to rising global temperatures (and the melted ice all goes into the lake), which ice chunk would give the greatest increase in the level of the lake water, if any?

Exercise:

Problem:

How do jogging on soft ground and wearing padded shoes reduce the pressures to which the feet and legs are subjected?

Exercise:

Problem:

Toe dancing (as in ballet) is much harder on toes than normal dancing or walking. Explain in terms of pressure.

Exercise:

Problem:

How do you convert pressure units like millimeters of mercury, centimeters of water, and inches of mercury into units like newtons per meter squared without resorting to a table of pressure conversion factors?

Problems & Exercises

Exercise:

Problem:

As a woman walks, her entire weight is momentarily placed on one heel of her high-heeled shoes. Calculate the pressure exerted on the floor by the heel if it has an area of $1.50\ \text{cm}^2$ and the woman's mass is 55.0 kg. Express the pressure in Pa. (In the early days of commercial flight, women were not allowed to wear high-heeled shoes because aircraft floors were too thin to withstand such large pressures.)

Solution:

$3.59 \times 10^6\ \text{Pa}$; or $521\ \text{lb/in}^2$

Exercise:

Problem:

The pressure exerted by a phonograph needle on a record is surprisingly large. If the equivalent of 1.00 g is supported by a needle, the tip of which is a circle 0.200 mm in radius, what pressure is exerted on the record in $\mathrm{N/m^2}$?

Exercise:

Problem:

Nail tips exert tremendous pressures when they are hit by hammers because they exert a large force over a small area. What force must be exerted on a nail with a circular tip of 1.00 mm diameter to create a pressure of $3.00 \times 10^9\ \mathrm{N/m^2}$?(This high pressure is possible because the hammer striking the nail is brought to rest in such a short distance.)

Solution:

$2.36 \times 10^3\ \mathrm{N}$

Glossary

pressure
: the force per unit area perpendicular to the force, over which the force acts

Variation of Pressure with Depth in a Fluid

- Define pressure in terms of weight.
- Explain the variation of pressure with depth in a fluid.
- Calculate density given pressure and altitude.

If your ears have ever popped on a plane flight or ached during a deep dive in a swimming pool, you have experienced the effect of depth on pressure in a fluid. At the Earth's surface, the air pressure exerted on you is a result of the weight of air above you. This pressure is reduced as you climb up in altitude and the weight of air above you decreases. Under water, the pressure exerted on you increases with increasing depth. In this case, the pressure being exerted upon you is a result of both the weight of water above you *and* that of the atmosphere above you. You may notice an air pressure change on an elevator ride that transports you many stories, but you need only dive a meter or so below the surface of a pool to feel a pressure increase. The difference is that water is much denser than air, about 775 times as dense.

Consider the container in [link]. Its bottom supports the weight of the fluid in it. Let us calculate the pressure exerted on the bottom by the weight of the fluid. That **pressure** is the weight of the fluid mg divided by the area A supporting it (the area of the bottom of the container):
Equation:

$$P = \frac{\text{mg}}{A}.$$

We can find the mass of the fluid from its volume and density:
Equation:

$$m = \rho V.$$

The volume of the fluid V is related to the dimensions of the container. It is
Equation:

$$V = \text{Ah},$$

where A is the cross-sectional area and h is the depth. Combining the last two equations gives

Equation:

$$m = \rho \mathrm{Ah}.$$

If we enter this into the expression for pressure, we obtain

Equation:

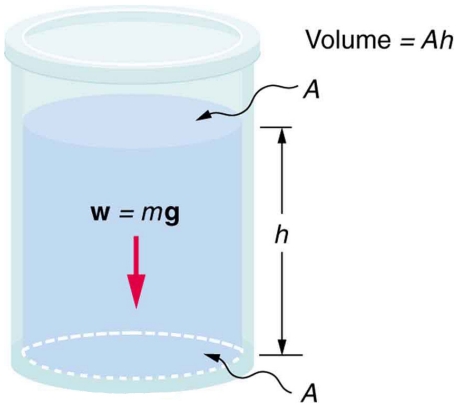
$$P = \frac{(\rho \mathrm{Ah})g}{A}.$$

The area cancels, and rearranging the variables yields

Equation:

$$P = h\rho g.$$

This value is the *pressure due to the weight of a fluid*. The equation has general validity beyond the special conditions under which it is derived here. Even if the container were not there, the surrounding fluid would still exert this pressure, keeping the fluid static. Thus the equation $P = h\rho g$ represents the pressure due to the weight of any fluid of *average density* ρ at any depth h below its surface. For liquids, which are nearly incompressible, this equation holds to great depths. For gases, which are quite compressible, one can apply this equation as long as the density changes are small over the depth considered. [link] illustrates this situation.



The bottom of this container supports the entire weight of the fluid in it. The vertical sides cannot exert an upward force on the fluid (since it cannot withstand a shearing force), and so the bottom must support it all.

Example:
Calculating the Average Pressure and Force Exerted: What Force Must a Dam Withstand?
In [link], we calculated the mass of water in a large reservoir. We will now consider the pressure and force acting on the dam retaining water. (See [link].) The dam is 500 m wide, and the water is 80.0 m deep at the dam. (a) What is the average pressure on the dam due to the water? (b) Calculate the force exerted against the dam and compare it with the weight of water in the dam (previously found to be 1.96×10^{13} N).
Strategy for (a)

The average pressure P due to the weight of the water is the pressure at the average depth h of 40.0 m, since pressure increases linearly with depth.

Solution for (a)

The average pressure due to the weight of a fluid is

Equation:

$$P = h\rho g.$$

Entering the density of water from [link] and taking h to be the average depth of 40.0 m, we obtain

Equation:

$$\begin{aligned} P &= (40.0\ \text{m})\left(10^3\ \tfrac{\text{kg}}{\text{m}^3}\right)\left(9.80\ \tfrac{\text{m}}{\text{s}^2}\right) \\ &= 3.92 \times 10^5\ \tfrac{\text{N}}{\text{m}^2} = 392\ \text{kPa}. \end{aligned}$$

Strategy for (b)

The force exerted on the dam by the water is the average pressure times the area of contact:

Equation:

$$F = PA.$$

Solution for (b)

We have already found the value for P. The area of the dam is $A = 80.0\ \text{m} \times 500\ \text{m} = 4.00 \times 10^4\ \text{m}^2$, so that

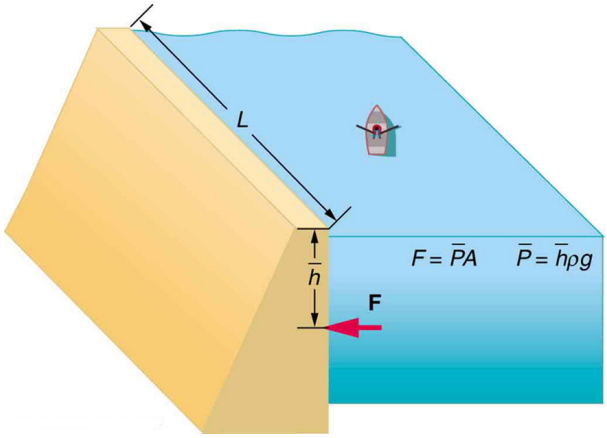
Equation:

$$\begin{aligned} F &= (3.92 \times 10^5\ \text{N/m}^2)(4.00 \times 10^4\ \text{m}^2) \\ &= 1.57 \times 10^{10}\ \text{N}. \end{aligned}$$

Discussion

Although this force seems large, it is small compared with the 1.96×10^{13} N weight of the water in the reservoir—in fact, it is only 0.0800% of the weight. Note that the pressure found in part (a) is completely independent of the width and length of the lake—it depends only on its average depth at the dam. Thus the force depends only on the

water’s average depth and the dimensions of the dam, *not* on the horizontal extent of the reservoir. In the diagram, the thickness of the dam increases with depth to balance the increasing force due to the increasing pressure.epth to balance the increasing force due to the increasing pressure.

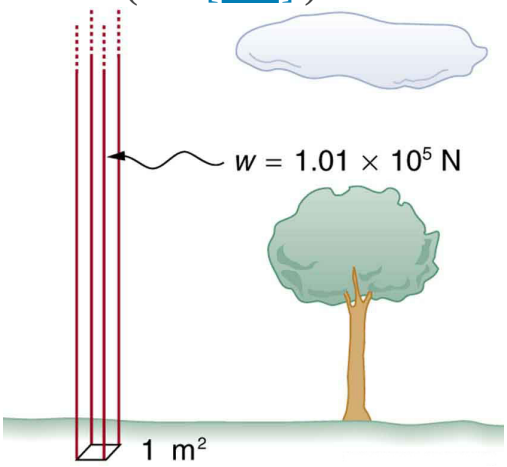


The dam must withstand the force exerted against it by the water it retains. This force is small compared with the weight of the water behind the dam.

Atmospheric pressure is another example of pressure due to the weight of a fluid, in this case due to the weight of *air* above a given height. The atmospheric pressure at the Earth’s surface varies a little due to the large-scale flow of the atmosphere induced by the Earth’s rotation (this creates weather “highs” and “lows”). However, the average pressure at sea level is given by the *standard atmospheric pressure* P_{atm}, measured to be
Equation:

$$1 \text{ atmosphere (atm)} = P_{\text{atm}} = 1.01 \times 10^5 \text{ N/m}^2 = 101 \text{ kPa}.$$

This relationship means that, on average, at sea level, a column of air above $1.00\ \text{m}^2$ of the Earth's surface has a weight of $1.01 \times 10^5\ \text{N}$, equivalent to 1 atm. (See [link].)



Atmospheric pressure at sea level averages $1.01 \times 10^5\ \text{Pa}$ (equivalent to 1 atm), since the column of air over this $1\ \text{m}^2$, extending to the top of the atmosphere, weighs $1.01 \times 10^5\ \text{N}$.

Example:
Calculating Average Density: How Dense Is the Air?
Calculate the average density of the atmosphere, given that it extends to an altitude of 120 km. Compare this density with that of air listed in [link].
Strategy
If we solve $P = h\rho g$ for density, we see that
Equation:

$$\rho = \frac{P}{\text{hg}}.$$

We then take P to be atmospheric pressure, h is given, and g is known, and so we can use this to calculate ρ.

Solution

Entering known values into the expression for ρ yields

Equation:

$$\rho = \frac{1.01 \times 10^5 \text{ N/m}^2}{(120 \times 10^3 \text{ m})(9.80 \text{ m/s}^2)} = 8.59 \times 10^{-2} \text{ kg/m}^3.$$

Discussion

This result is the average density of air between the Earth's surface and the top of the Earth's atmosphere, which essentially ends at 120 km. The density of air at sea level is given in [link] as 1.29 kg/m^3 —about 15 times its average value. Because air is so compressible, its density has its highest value near the Earth's surface and declines rapidly with altitude.

Example:

Calculating Depth Below the Surface of Water: What Depth of Water Creates the Same Pressure as the Entire Atmosphere?

Calculate the depth below the surface of water at which the pressure due to the weight of the water equals 1.00 atm.

Strategy

We begin by solving the equation $P = h\rho g$ for depth h:

Equation:

$$h = \frac{P}{\rho g}.$$

Then we take P to be 1.00 atm and ρ to be the density of the water that creates the pressure.

Solution

Entering the known values into the expression for h gives

Equation:

$$h = \frac{1.01 \times 10^5 \text{ N/m}^2}{(1.00 \times 10^3 \text{ kg/m}^3)(9.80 \text{ m/s}^2)} = 10.3 \text{ m}.$$

Discussion
Just 10.3 m of water creates the same pressure as 120 km of air. Since water is nearly incompressible, we can neglect any change in its density over this depth.

What do you suppose is the *total* pressure at a depth of 10.3 m in a swimming pool? Does the atmospheric pressure on the water's surface affect the pressure below? The answer is yes. This seems only logical, since both the water's weight and the atmosphere's weight must be supported. So the *total* pressure at a depth of 10.3 m is 2 atm—half from the water above and half from the air above. We shall see in Pascal's Principle that fluid pressures always add in this way.

Section Summary

- Pressure is the weight of the fluid mg divided by the area A supporting it (the area of the bottom of the container):
  Equation:

$$P = \frac{\text{mg}}{A}.$$

- Pressure due to the weight of a liquid is given by
  Equation:

$$P = h\rho g,$$

  where P is the pressure, h is the height of the liquid, ρ is the density of the liquid, and g is the acceleration due to gravity.

Conceptual Questions

Exercise:

Problem:

Atmospheric pressure exerts a large force (equal to the weight of the atmosphere above your body—about 10 tons) on the top of your body when you are lying on the beach sunbathing. Why are you able to get up?

Exercise:

Problem:

Why does atmospheric pressure decrease more rapidly than linearly with altitude?

Exercise:

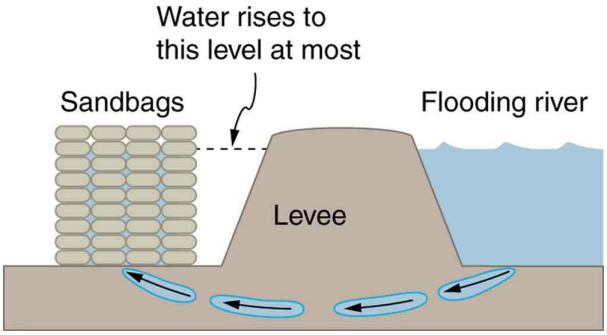
Problem:

What are two reasons why mercury rather than water is used in barometers?

Exercise:

Problem:

[link] shows how sandbags placed around a leak outside a river levee can effectively stop the flow of water under the levee. Explain how the small amount of water inside the column formed by the sandbags is able to balance the much larger body of water behind the levee.



Because the river level is very high, it has started to leak under the levee. Sandbags are placed around the leak, and the water held by them rises until it is the same level as the river, at which point the water there stops rising.

Exercise:

Problem:

Why is it difficult to swim under water in the Great Salt Lake?

Exercise:

Problem:

Is there a net force on a dam due to atmospheric pressure? Explain your answer.

Exercise:

Problem:

Does atmospheric pressure add to the gas pressure in a rigid tank? In a toy balloon? When, in general, does atmospheric pressure *not* affect the total pressure in a fluid?

Exercise:

Problem:

You can break a strong wine bottle by pounding a cork into it with your fist, but the cork must press directly against the liquid filling the bottle—there can be no air between the cork and liquid. Explain why the bottle breaks, and why it will not if there is air between the cork and liquid.

Problems & Exercises

Exercise:

Problem: What depth of mercury creates a pressure of 1.00 atm?

Solution:

0.760 m

Exercise:

Problem:

The greatest ocean depths on the Earth are found in the Marianas Trench near the Philippines. Calculate the pressure due to the ocean at the bottom of this trench, given its depth is 11.0 km and assuming the density of seawater is constant all the way down.

Exercise:

Problem: Verify that the SI unit of $h\rho g$ is $\mathrm{N/m^2}$.

Solution:
Equation:

$$\begin{aligned}(h\rho g)_{\text{units}} &= (\mathrm{m})\left(\mathrm{kg/m^3}\right)\left(\mathrm{m/s^2}\right) = \left(\mathrm{kg \cdot m^2}\right)/\left(\mathrm{m^3 \cdot s^2}\right) \\ &= \left(\mathrm{kg \cdot m/s^2}\right)\left(1/\mathrm{m^2}\right) \\ &= \mathrm{N/m^2}\end{aligned}$$

Exercise:

Problem:

Water towers store water above the level of consumers for times of heavy use, eliminating the need for high-speed pumps. How high above a user must the water level be to create a gauge pressure of $3.00 \times 10^5 \text{ N/m}^2$?

Exercise:

Problem:

The aqueous humor in a person's eye is exerting a force of 0.300 N on the 1.10-cm^2 area of the cornea. (a) What pressure is this in mm Hg? (b) Is this value within the normal range for pressures in the eye?

Solution:

(a) 20.5 mm Hg

(b) The range of pressures in the eye is 12–24 mm Hg, so the result in part (a) is within that range

Exercise:

Problem:

How much force is exerted on one side of an 8.50 cm by 11.0 cm sheet of paper by the atmosphere? How can the paper withstand such a force?

Exercise:

Problem:

What pressure is exerted on the bottom of a 0.500-m-wide by 0.900-m-long gas tank that can hold 50.0 kg of gasoline by the weight of the gasoline in it when it is full?

Solution:

$1.09 \times 10^3 \text{ N/m}^2$

Exercise:

Problem:

Calculate the average pressure exerted on the palm of a shot-putter's hand by the shot if the area of contact is $50.0\ \mathrm{cm}^2$ and he exerts a force of 800 N on it. Express the pressure in $\mathrm{N/m}^2$ and compare it with the $1.00 \times 10^6\ \mathrm{Pa}$ pressures sometimes encountered in the skeletal system.

Exercise:

Problem:

The left side of the heart creates a pressure of 120 mm Hg by exerting a force directly on the blood over an effective area of $15.0\ \mathrm{cm}^2$. What force does it exert to accomplish this?

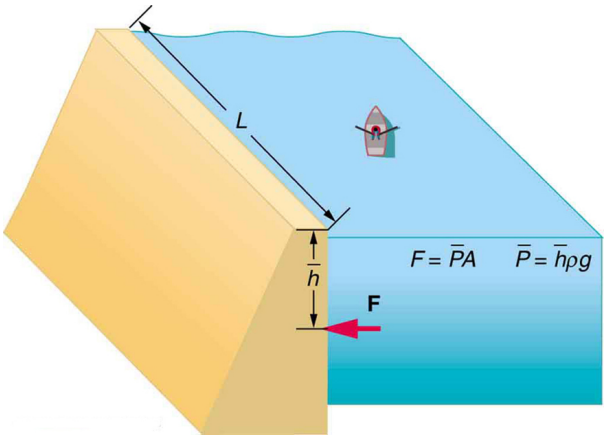
Solution:

24.0 N

Exercise:

Problem:

Show that the total force on a rectangular dam due to the water behind it increases with the *square* of the water depth. In particular, show that this force is given by $F = \rho g h^2 L/2$, where ρ is the density of water, h is its depth at the dam, and L is the length of the dam. You may assume the face of the dam is vertical. (Hint: Calculate the average pressure exerted and multiply this by the area in contact with the water. (See [link].)



Glossary

pressure
: the weight of the fluid divided by the area supporting it

Pascal’s Principle

- Define pressure.
- State Pascal’s principle.
- Understand applications of Pascal’s principle.
- Derive relationships between forces in a hydraulic system.

Pressure is defined as force per unit area. Can pressure be increased in a fluid by pushing directly on the fluid? Yes, but it is much easier if the fluid is enclosed. The heart, for example, increases blood pressure by pushing directly on the blood in an enclosed system (valves closed in a chamber). If you try to push on a fluid in an open system, such as a river, the fluid flows away. An enclosed fluid cannot flow away, and so pressure is more easily increased by an applied force.

What happens to a pressure in an enclosed fluid? Since atoms in a fluid are free to move about, they transmit the pressure to all parts of the fluid and to the walls of the container. Remarkably, the pressure is transmitted *undiminished*. This phenomenon is called **Pascal’s principle**, because it was first clearly stated by the French philosopher and scientist Blaise Pascal (1623–1662): A change in pressure applied to an enclosed fluid is transmitted undiminished to all portions of the fluid and to the walls of its container.

Note:
Pascal’s Principle
A change in pressure applied to an enclosed fluid is transmitted undiminished to all portions of the fluid and to the walls of its container.

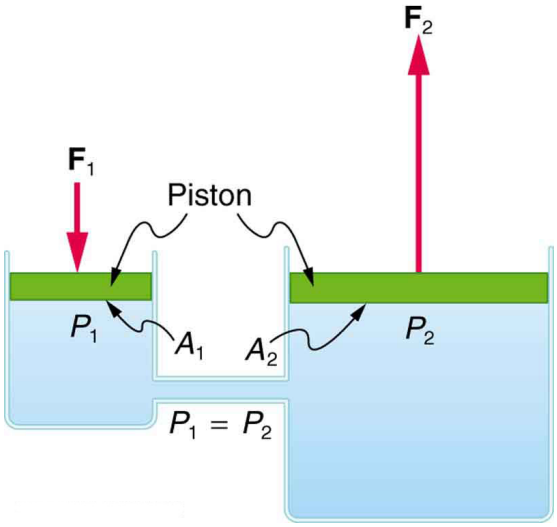
Pascal’s principle, an experimentally verified fact, is what makes pressure so important in fluids. Since a change in pressure is transmitted undiminished in an enclosed fluid, we often know more about pressure than other physical quantities in fluids. Moreover, Pascal’s principle implies that

the total pressure in a fluid is the sum of the pressures from different sources. We shall find this fact—that pressures add—very useful.

Blaise Pascal had an interesting life in that he was home-schooled by his father who removed all of the mathematics textbooks from his house and forbade him to study mathematics until the age of 15. This, of course, raised the boy's curiosity, and by the age of 12, he started to teach himself geometry. Despite this early deprivation, Pascal went on to make major contributions in the mathematical fields of probability theory, number theory, and geometry. He is also well known for being the inventor of the first mechanical digital calculator, in addition to his contributions in the field of fluid statics.

Application of Pascal's Principle

One of the most important technological applications of Pascal's principle is found in a *hydraulic system*, which is an enclosed fluid system used to exert forces. The most common hydraulic systems are those that operate car brakes. Let us first consider the simple hydraulic system shown in [link].



A typical hydraulic system with two fluid-filled cylinders, capped with

pistons and connected by a tube called a hydraulic line. A downward force $\mathbf{F}_1$ on the left piston creates a pressure that is transmitted undiminished to all parts of the enclosed fluid. This results in an upward force $\mathbf{F}_2$ on the right piston that is larger than $\mathbf{F}_1$ because the right piston has a larger area.

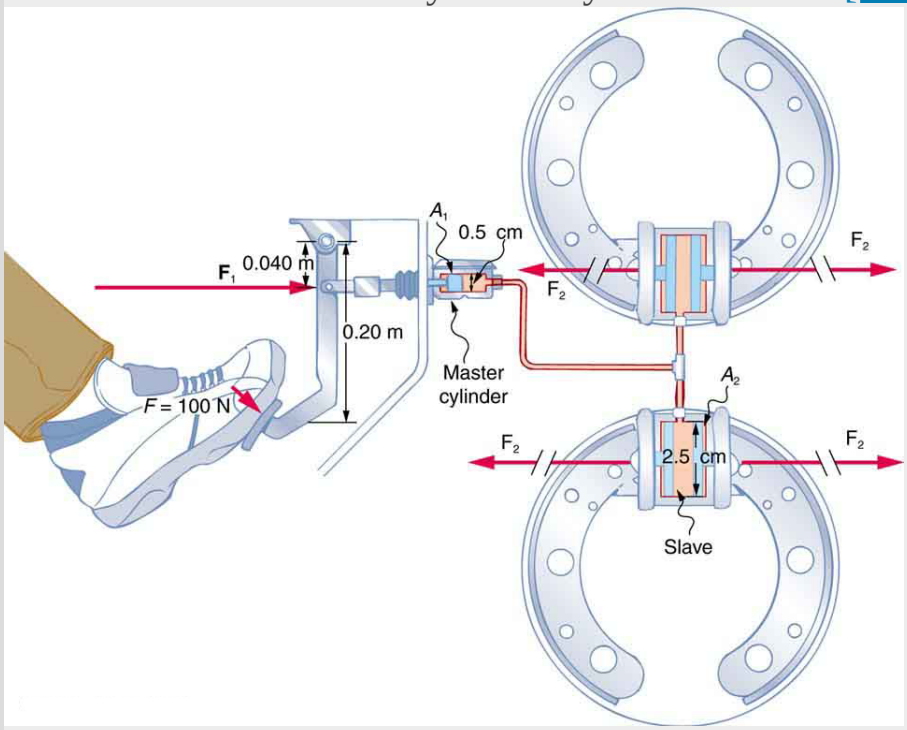
Relationship Between Forces in a Hydraulic System

We can derive a relationship between the forces in the simple hydraulic system shown in [link] by applying Pascal's principle. Note first that the two pistons in the system are at the same height, and so there will be no difference in pressure due to a difference in depth. Now the pressure due to F_1 acting on area A_1 is simply $P_1 = \frac{F_1}{A_1}$, as defined by $P = \frac{F}{A}$. According to Pascal's principle, this pressure is transmitted undiminished throughout the fluid and to all walls of the container. Thus, a pressure P_2 is felt at the other piston that is equal to P_1. That is $P_1 = P_2$.

But since $P_2 = \frac{F_2}{A_2}$, we see that $\frac{F_1}{A_1} = \frac{F_2}{A_2}$.

This equation relates the ratios of force to area in any hydraulic system, providing the pistons are at the same vertical height and that friction in the system is negligible. Hydraulic systems can increase or decrease the force applied to them. To make the force larger, the pressure is applied to a larger area. For example, if a 100-N force is applied to the left cylinder in [link] and the right one has an area five times greater, then the force out is 500 N. Hydraulic systems are analogous to simple levers, but they have the advantage that pressure can be sent through tortuously curved lines to several places at once.

Example:
Calculating Force of Slave Cylinders: Pascal Puts on the Brakes
Consider the automobile hydraulic system shown in [link].



Hydraulic brakes use Pascal's principle. The driver exerts a force of 100 N on the brake pedal. This force is increased by the simple lever and again by the hydraulic system. Each of the identical slave cylinders receives the same pressure and, therefore, creates the same force output F_2. The circular cross-sectional areas of the master and slave cylinders are represented by A_1 and A_2, respectively

A force of 100 N is applied to the brake pedal, which acts on the cylinder—called the master—through a lever. A force of 500 N is exerted on the master cylinder. (The reader can verify that the force is 500 N using techniques of statics from Applications of Statics, Including Problem-Solving Strategies.) Pressure created in the master cylinder is transmitted to four so-called slave cylinders. The master cylinder has a diameter of

0.500 cm, and each slave cylinder has a diameter of 2.50 cm. Calculate the force F_2 created at each of the slave cylinders.

Strategy

We are given the force F_1 that is applied to the master cylinder. The cross-sectional areas A_1 and A_2 can be calculated from their given diameters. Then $\frac{F_1}{A_1} = \frac{F_2}{A_2}$ can be used to find the force F_2. Manipulate this algebraically to get F_2 on one side and substitute known values:

Solution

Pascal's principle applied to hydraulic systems is given by $\frac{F_1}{A_1} = \frac{F_2}{A_2}$:

Equation:

$$F_2 = \frac{A_2}{A_1}F_1 = \frac{\pi r_2^2}{\pi r_1^2}F_1 = \frac{(1.25\text{ cm})^2}{(0.250\text{ cm})^2} \times 500\text{ N} = 1.25 \times 10^4\text{ N}.$$

Discussion

This value is the force exerted by each of the four slave cylinders. Note that we can add as many slave cylinders as we wish. If each has a 2.50-cm diameter, each will exert 1.25×10^4 N.

A simple hydraulic system, such as a simple machine, can increase force but cannot do more work than done on it. Work is force times distance moved, and the slave cylinder moves through a smaller distance than the master cylinder. Furthermore, the more slaves added, the smaller the distance each moves. Many hydraulic systems—such as power brakes and those in bulldozers—have a motorized pump that actually does most of the work in the system. The movement of the legs of a spider is achieved partly by hydraulics. Using hydraulics, a jumping spider can create a force that makes it capable of jumping 25 times its length!

Note:

Making Connections: Conservation of Energy

Conservation of energy applied to a hydraulic system tells us that the system cannot do more work than is done on it. Work transfers energy, and so the work output cannot exceed the work input. Power brakes and other similar hydraulic systems use pumps to supply extra energy when needed.

Section Summary

- Pressure is force per unit area.
- A change in pressure applied to an enclosed fluid is transmitted undiminished to all portions of the fluid and to the walls of its container.
- A hydraulic system is an enclosed fluid system used to exert forces.

Conceptual Questions

Exercise:

Problem:

Suppose the master cylinder in a hydraulic system is at a greater height than the slave cylinder. Explain how this will affect the force produced at the slave cylinder.

Problems & Exercises

Exercise:

Problem:

How much pressure is transmitted in the hydraulic system considered in [link]? Express your answer in pascals and in atmospheres.

Solution:

2.55×10^7 Pa; or 251 atm

Exercise:

Problem:

What force must be exerted on the master cylinder of a hydraulic lift to support the weight of a 2000-kg car (a large car) resting on the slave cylinder? The master cylinder has a 2.00-cm diameter and the slave has a 24.0-cm diameter.

Exercise:

Problem:

A crass host pours the remnants of several bottles of wine into a jug after a party. He then inserts a cork with a 2.00-cm diameter into the bottle, placing it in direct contact with the wine. He is amazed when he pounds the cork into place and the bottom of the jug (with a 14.0-cm diameter) breaks away. Calculate the extra force exerted against the bottom if he pounded the cork with a 120-N force.

Solution:

5.76×10^3 N extra force

Exercise:

Problem:

A certain hydraulic system is designed to exert a force 100 times as large as the one put into it. (a) What must be the ratio of the area of the slave cylinder to the area of the master cylinder? (b) What must be the ratio of their diameters? (c) By what factor is the distance through which the output force moves reduced relative to the distance through which the input force moves? Assume no losses to friction.

Exercise:

Problem:

(a) Verify that work input equals work output for a hydraulic system assuming no losses to friction. Do this by showing that the distance the output force moves is reduced by the same factor that the output force is increased. Assume the volume of the fluid is constant. (b) What effect would friction within the fluid and between components in the system have on the output force? How would this depend on whether or not the fluid is moving?

Solution:

(a) $V = d_{\mathrm{i}}A_{\mathrm{i}} = d_{\mathrm{o}}A_{\mathrm{o}} \Rightarrow d_{\mathrm{o}} = d_{\mathrm{i}}\left(\frac{A_{\mathrm{i}}}{A_{\mathrm{o}}}\right)$.

Now, using equation:
Equation:

$$\frac{F_1}{A_1} = \frac{F_2}{A_2} \Rightarrow F_{\mathrm{o}} = F_{\mathrm{i}} \quad \frac{A_{\mathrm{o}}}{A_{\mathrm{i}}} \quad .$$

Finally,
Equation:

$$W_{\mathrm{o}} = F_{\mathrm{o}}d_{\mathrm{o}} = \quad \frac{F_{\mathrm{i}}A_{\mathrm{o}}}{A_{\mathrm{i}}} \quad \frac{d_{\mathrm{i}}A_{\mathrm{i}}}{A_{\mathrm{o}}} \quad = F_{\mathrm{i}}d_{\mathrm{i}} = W_{\mathrm{i}}.$$

In other words, the work output equals the work input.

(b) If the system is not moving, friction would not play a role. With friction, we know there are losses, so that $W_{\mathrm{out}} = W_{\mathrm{in}} - W_{\mathrm{f}}$; therefore, the work output is less than the work input. In other words, with friction, you need to push harder on the input piston than was calculated for the nonfriction case.

Glossary

Pascal’s Principle
: a change in pressure applied to an enclosed fluid is transmitted undiminished to all portions of the fluid and to the walls of its container

Gauge Pressure, Absolute Pressure, and Pressure Measurement

- Define gauge pressure and absolute pressure.
- Understand the working of aneroid and open-tube barometers.

If you limp into a gas station with a nearly flat tire, you will notice the tire gauge on the airline reads nearly zero when you begin to fill it. In fact, if there were a gaping hole in your tire, the gauge would read zero, even though atmospheric pressure exists in the tire. Why does the gauge read zero? There is no mystery here. Tire gauges are simply designed to read zero at atmospheric pressure and positive when pressure is greater than atmospheric.

Similarly, atmospheric pressure adds to blood pressure in every part of the circulatory system. (As noted in Pascal's Principle, the total pressure in a fluid is the sum of the pressures from different sources—here, the heart and the atmosphere.) But atmospheric pressure has no net effect on blood flow since it adds to the pressure coming out of the heart and going back into it, too. What is important is how much *greater* blood pressure is than atmospheric pressure. Blood pressure measurements, like tire pressures, are thus made relative to atmospheric pressure.

In brief, it is very common for pressure gauges to ignore atmospheric pressure—that is, to read zero at atmospheric pressure. We therefore define **gauge pressure** to be the pressure relative to atmospheric pressure. Gauge pressure is positive for pressures above atmospheric pressure, and negative for pressures below it.

Note:
Gauge Pressure
Gauge pressure is the pressure relative to atmospheric pressure. Gauge pressure is positive for pressures above atmospheric pressure, and negative for pressures below it.

In fact, atmospheric pressure does add to the pressure in any fluid not enclosed in a rigid container. This happens because of Pascal's principle. The total pressure, or **absolute pressure**, is thus the sum of gauge pressure and atmospheric pressure: $P_{\text{abs}} = P_{\text{g}} + P_{\text{atm}}$ where P_{abs} is absolute pressure, P_{g} is gauge pressure, and P_{atm} is atmospheric pressure. For example, if your tire gauge reads 34 psi

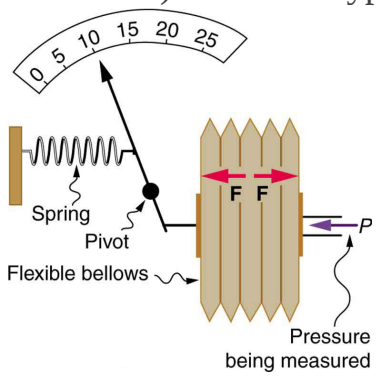
(pounds per square inch), then the absolute pressure is 34 psi plus 14.7 psi (P_{atm} in psi), or 48.7 psi (equivalent to 336 kPa).

Note:
Absolute Pressure
Absolute pressure is the sum of gauge pressure and atmospheric pressure.

For reasons we will explore later, in most cases the absolute pressure in fluids cannot be negative. Fluids push rather than pull, so the smallest absolute pressure is zero. (A negative absolute pressure is a pull.) Thus the smallest possible gauge pressure is $P_{\text{g}} = -P_{\text{atm}}$ (this makes P_{abs} zero). There is no theoretical limit to how large a gauge pressure can be.

There are a host of devices for measuring pressure, ranging from tire gauges to blood pressure cuffs. Pascal's principle is of major importance in these devices. The undiminished transmission of pressure through a fluid allows precise remote sensing of pressures. Remote sensing is often more convenient than putting a measuring device into a system, such as a person's artery.

[link] shows one of the many types of mechanical pressure gauges in use today. In all mechanical pressure gauges, pressure results in a force that is converted (or transduced) into some type of readout.

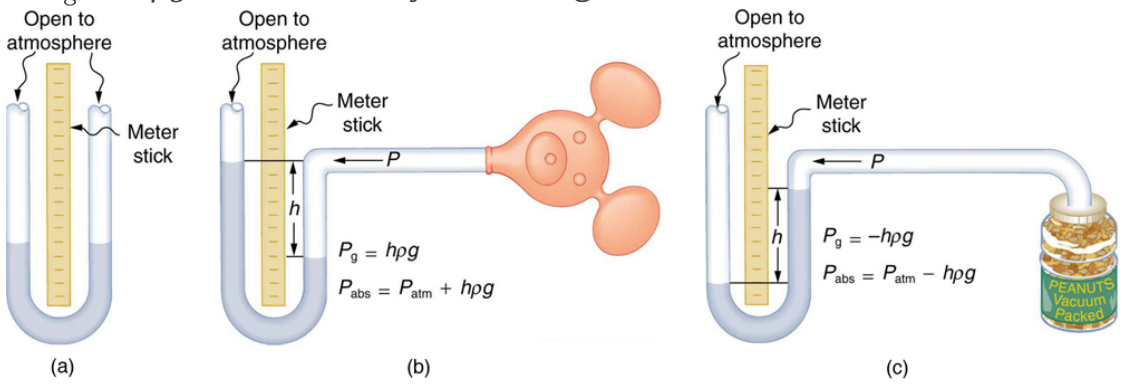


This aneroid gauge utilizes flexible bellows connected to a mechanical

indicator to measure pressure.

An entire class of gauges uses the property that pressure due to the weight of a fluid is given by $P = h\rho g$. Consider the U-shaped tube shown in [link], for example. This simple tube is called a *manometer*. In [link](a), both sides of the tube are open to the atmosphere. Atmospheric pressure therefore pushes down on each side equally so its effect cancels. If the fluid is deeper on one side, there is a greater pressure on the deeper side, and the fluid flows away from that side until the depths are equal.

Let us examine how a manometer is used to measure pressure. Suppose one side of the U-tube is connected to some source of pressure P_{abs} such as the toy balloon in [link](b) or the vacuum-packed peanut jar shown in [link](c). Pressure is transmitted undiminished to the manometer, and the fluid levels are no longer equal. In [link](b), P_{abs} is greater than atmospheric pressure, whereas in [link](c), P_{abs} is less than atmospheric pressure. In both cases, P_{abs} differs from atmospheric pressure by an amount $h\rho g$, where ρ is the density of the fluid in the manometer. In [link](b), P_{abs} can support a column of fluid of height h, and so it must exert a pressure $h\rho g$ greater than atmospheric pressure (the gauge pressure P_{g} is positive). In [link](c), atmospheric pressure can support a column of fluid of height h, and so P_{abs} is less than atmospheric pressure by an amount $h\rho g$ (the gauge pressure P_{g} is negative). A manometer with one side open to the atmosphere is an ideal device for measuring gauge pressures. The gauge pressure is $P_{\text{g}} = h\rho g$ and is found by measuring h.



An open-tube manometer has one side open to the atmosphere. (a) Fluid depth must be the same on both sides, or the pressure each side exerts at the bottom will be unequal and there will be flow from the

deeper side. (b) A positive gauge pressure $P_{\text{g}} = h\rho g$ transmitted to one side of the manometer can support a column of fluid of height h. (c) Similarly, atmospheric pressure is greater than a negative gauge pressure P_{g} by an amount $h\rho g$. The jar's rigidity prevents atmospheric pressure from being transmitted to the peanuts.

Mercury manometers are often used to measure arterial blood pressure. An inflatable cuff is placed on the upper arm as shown in [link]. By squeezing the bulb, the person making the measurement exerts pressure, which is transmitted undiminished to both the main artery in the arm and the manometer. When this applied pressure exceeds blood pressure, blood flow below the cuff is cut off. The person making the measurement then slowly lowers the applied pressure and listens for blood flow to resume. Blood pressure pulsates because of the pumping action of the heart, reaching a maximum, called **systolic pressure**, and a minimum, called **diastolic pressure**, with each heartbeat. Systolic pressure is measured by noting the value of h when blood flow first begins as cuff pressure is lowered. Diastolic pressure is measured by noting h when blood flows without interruption. The typical blood pressure of a young adult raises the mercury to a height of 120 mm at systolic and 80 mm at diastolic. This is commonly quoted as 120 over 80, or 120/80. The first pressure is representative of the maximum output of the heart; the second is due to the elasticity of the arteries in maintaining the pressure between beats. The density of the mercury fluid in the manometer is 13.6 times greater than water, so the height of the fluid will be 1/13.6 of that in a water manometer. This reduced height can make measurements difficult, so mercury manometers are used to measure larger pressures, such as blood pressure. The density of mercury is such that $1.0 \text{ mm Hg} = 133 \text{ Pa}$.

Note:
Systolic Pressure
Systolic pressure is the maximum blood pressure.

Note:
Diastolic Pressure
Diastolic pressure is the minimum blood pressure.



In routine blood pressure measurements, an inflatable cuff is placed on the upper arm at the same level as the heart. Blood flow is detected just below the cuff, and corresponding pressures are transmitted to a mercury-filled manometer. (credit: U.S. Army photo by Spc. Micah E. Clare\4TH BCT)

Example:
Calculating Height of IV Bag: Blood Pressure and Intravenous Infusions

Intravenous infusions are usually made with the help of the gravitational force. Assuming that the density of the fluid being administered is 1.00 g/ml, at what height should the IV bag be placed above the entry point so that the fluid just enters the vein if the blood pressure in the vein is 18 mm Hg above atmospheric pressure? Assume that the IV bag is collapsible.

Strategy for (a)

For the fluid to just enter the vein, its pressure at entry must exceed the blood pressure in the vein (18 mm Hg above atmospheric pressure). We therefore need to find the height of fluid that corresponds to this gauge pressure.

Solution

We first need to convert the pressure into SI units. Since $1.0 \text{ mm Hg} = 133 \text{ Pa}$,

Equation:

$$P = 18 \text{ mm Hg} \times \frac{133 \text{ Pa}}{1.0 \text{ mm Hg}} = 2400 \text{ Pa}.$$

Rearranging $P_{\text{g}} = h\rho g$ for h gives $h = \frac{P_{\text{g}}}{\rho g}$. Substituting known values into this equation gives

Equation:

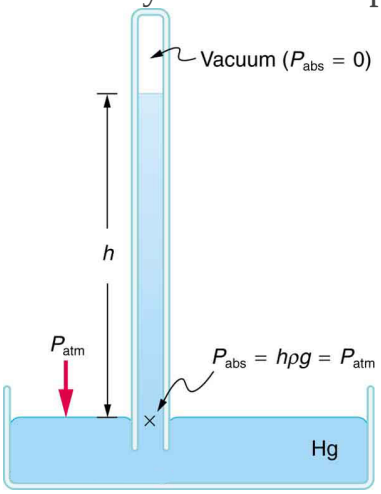
$$\begin{aligned} h &= \frac{2400 \text{ N/m}^2}{\left(1.0 \times 10^3 \text{ kg/m}^3\right)\left(9.80 \text{ m/s}^2\right)} \\ &= 0.24 \text{ m}. \end{aligned}$$

Discussion

The IV bag must be placed at 0.24 m above the entry point into the arm for the fluid to just enter the arm. Generally, IV bags are placed higher than this. You may have noticed that the bags used for blood collection are placed below the donor to allow blood to flow easily from the arm to the bag, which is the opposite direction of flow than required in the example presented here.

A *barometer* is a device that measures atmospheric pressure. A mercury barometer is shown in [link]. This device measures atmospheric pressure, rather than gauge pressure, because there is a nearly pure vacuum above the mercury in the tube. The height of the mercury is such that $h\rho g = P_{\text{atm}}$. When atmospheric pressure varies, the mercury rises or falls, giving important clues to weather forecasters. The barometer can also be used as an altimeter, since average atmospheric pressure varies with altitude. Mercury barometers and manometers

are so common that units of mm Hg are often quoted for atmospheric pressure and blood pressures. [link] gives conversion factors for some of the more commonly used units of pressure.



A mercury barometer measures atmospheric pressure. The pressure due to the mercury's weight, $h\rho g$, equals atmospheric pressure. The atmosphere is able to force mercury in the tube to a height h because the pressure above the mercury is zero.

Conversion to $\mathrm{N/m^2}$ (Pa)	Conversion from atm
$1.0\ \mathrm{atm} = 1.013 \times 10^5\ \mathrm{N/m^2}$	$1.0\ \mathrm{atm} = 1.013 \times 10^5\ \mathrm{N/m^2}$
$1.0\ \mathrm{dyne/cm^2} = 0.10\ \mathrm{N/m^2}$	$1.0\ \mathrm{atm} = 1.013 \times 10^6\ \mathrm{dyne/cm^2}$
$1.0\ \mathrm{kg/cm^2} = 9.8 \times 10^4\ \mathrm{N/m^2}$	$1.0\ \mathrm{atm} = 1.013\ \mathrm{kg/cm^2}$
$1.0\ \mathrm{lb/in.^2} = 6.90 \times 10^3\ \mathrm{N/m^2}$	$1.0\ \mathrm{atm} = 14.7\ \mathrm{lb/in.^2}$
$1.0\ \mathrm{mm\ Hg} = 133\ \mathrm{N/m^2}$	$1.0\ \mathrm{atm} = 760\ \mathrm{mm\ Hg}$
$1.0\ \mathrm{cm\ Hg} = 1.33 \times 10^3\ \mathrm{N/m^2}$	$1.0\ \mathrm{atm} = 76.0\ \mathrm{cm\ Hg}$
$1.0\ \mathrm{cm\ water} = 98.1\ \mathrm{N/m^2}$	$1.0\ \mathrm{atm} = 1.03 \times 10^3\ \mathrm{cm\ water}$
$1.0\ \mathrm{bar} = 1.000 \times 10^5\ \mathrm{N/m^2}$	$1.0\ \mathrm{atm} = 1.013\ \mathrm{bar}$
$1.0\ \mathrm{millibar} = 1.000 \times 10^2\ \mathrm{N/m^2}$	$1.0\ \mathrm{atm} = 1013\ \mathrm{millibar}$

Conversion Factors for Various Pressure Units

Section Summary

- Gauge pressure is the pressure relative to atmospheric pressure.
- Absolute pressure is the sum of gauge pressure and atmospheric pressure.
- Aneroid gauge measures pressure using a bellows-and-spring arrangement connected to the pointer of a calibrated scale.
- Open-tube manometers have U-shaped tubes and one end is always open. It is used to measure pressure.
- A mercury barometer is a device that measures atmospheric pressure.

Conceptual Questions

Exercise:

Problem:

Explain why the fluid reaches equal levels on either side of a manometer if both sides are open to the atmosphere, even if the tubes are of different diameters.

Exercise:

Problem:

[link] shows how a common measurement of arterial blood pressure is made. Is there any effect on the measured pressure if the manometer is lowered? What is the effect of raising the arm above the shoulder? What is the effect of placing the cuff on the upper leg with the person standing? Explain your answers in terms of pressure created by the weight of a fluid.

Exercise:

Problem:

Considering the magnitude of typical arterial blood pressures, why are mercury rather than water manometers used for these measurements?

Problems & Exercises

Exercise:

Problem:

Find the gauge and absolute pressures in the balloon and peanut jar shown in [link], assuming the manometer connected to the balloon uses water whereas the manometer connected to the jar contains mercury. Express in units of centimeters of water for the balloon and millimeters of mercury for the jar, taking $h = 0.0500$ m for each.

Solution:

Balloon:

$$\begin{aligned} P_{\text{g}} &= 5.00 \text{ cm } \text{H}_2\text{O}, \\ P_{\text{abs}} &= 1.035 \times 10^3 \text{ cm } \text{H}_2\text{O}. \end{aligned}$$

Jar:

$$\begin{aligned} P_{\text{g}} &= -50.0 \text{ mm Hg}, \\ P_{\text{abs}} &= 710 \text{ mm Hg}. \end{aligned}$$

Exercise:

Problem:

(a) Convert normal blood pressure readings of 120 over 80 mm Hg to newtons per meter squared using the relationship for pressure due to the weight of a fluid $(P = h\rho g)$ rather than a conversion factor. (b) Discuss why blood pressures for an infant could be smaller than those for an adult. Specifically, consider the smaller height to which blood must be pumped.

Exercise:

Problem:

How tall must a water-filled manometer be to measure blood pressures as high as 300 mm Hg?

Solution:

4.08 m

Exercise:

Problem:

Pressure cookers have been around for more than 300 years, although their use has strongly declined in recent years (early models had a nasty habit of exploding). How much force must the latches holding the lid onto a pressure cooker be able to withstand if the circular lid is $25.0\ \text{cm}$ in diameter and the gauge pressure inside is 300 atm? Neglect the weight of the lid.

Exercise:

Problem:

Suppose you measure a standing person's blood pressure by placing the cuff on his leg 0.500 m below the heart. Calculate the pressure you would observe (in units of mm Hg) if the pressure at the heart were 120 over 80 mm Hg. Assume that there is no loss of pressure due to resistance in the circulatory system (a reasonable assumption, since major arteries are large).

Solution:

$$\Delta P = 38.7\ \text{mm Hg},$$

$$\text{Leg blood pressure} = \frac{159}{119}.$$

Exercise:

Problem:

A submarine is stranded on the bottom of the ocean with its hatch 25.0 m below the surface. Calculate the force needed to open the hatch from the inside, given it is circular and 0.450 m in diameter. Air pressure inside the submarine is 1.00 atm.

Exercise:

Problem:

Assuming bicycle tires are perfectly flexible and support the weight of bicycle and rider by pressure alone, calculate the total area of the tires in contact with the ground. The bicycle plus rider has a mass of 80.0 kg, and the gauge pressure in the tires is $3.50 \times 10^5\ \text{Pa}$.

Solution:

$22.4\ \text{cm}^2$

Glossary

absolute pressure
: the sum of gauge pressure and atmospheric pressure

diastolic pressure
: the minimum blood pressure in the artery

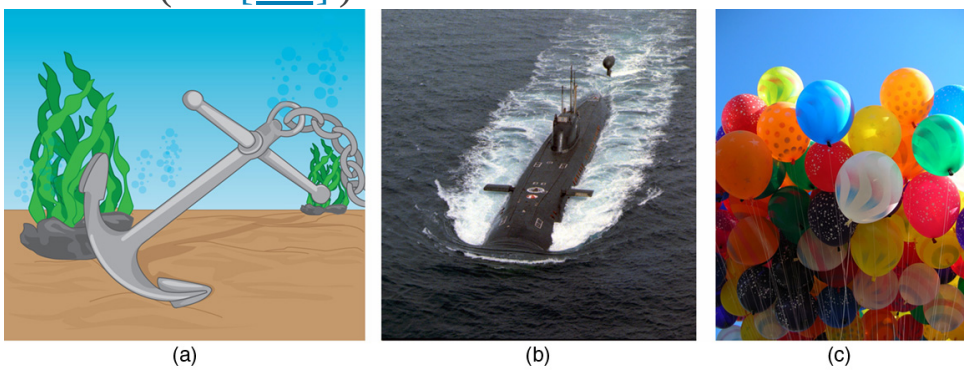
gauge pressure
: the pressure relative to atmospheric pressure

systolic pressure
: the maximum blood pressure in the artery

Archimedes’ Principle

- Define buoyant force.
- State Archimedes’ principle.
- Understand why objects float or sink.
- Understand the relationship between density and Archimedes’ principle.

When you rise from lounging in a warm bath, your arms feel strangely heavy. This is because you no longer have the buoyant support of the water. Where does this buoyant force come from? Why is it that some things float and others do not? Do objects that sink get any support at all from the fluid? Is your body buoyed by the atmosphere, or are only helium balloons affected? (See [link].)

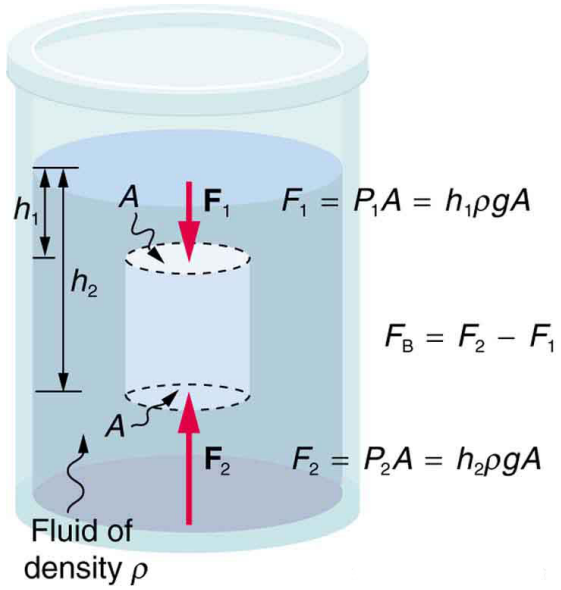

(a) (b) (c)

(a) Even objects that sink, like this anchor, are partly supported by water when submerged. (b) Submarines have adjustable density (ballast tanks) so that they may float or sink as desired. (credit: Allied Navy) (c) Helium-filled balloons tug upward on their strings, demonstrating air’s buoyant effect. (credit: Crystl)

Answers to all these questions, and many others, are based on the fact that pressure increases with depth in a fluid. This means that the upward force on the bottom of an object in a fluid is greater than the downward force on the top of the object. There is a net upward, or **buoyant force** on any object in any fluid. (See [link].) If the buoyant force is greater than the object’s

weight, the object will rise to the surface and float. If the buoyant force is less than the object's weight, the object will sink. If the buoyant force equals the object's weight, the object will remain suspended at that depth. The buoyant force is always present whether the object floats, sinks, or is suspended in a fluid.

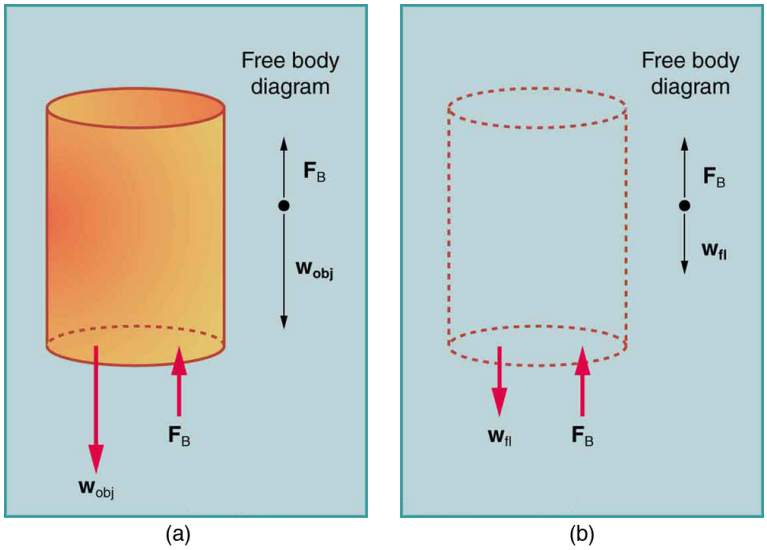
Note:
Buoyant Force
The buoyant force is the net upward force on any object in any fluid.



Pressure due to the weight of a fluid increases with depth since $P = h\rho g$. This pressure and associated upward force on the bottom of the cylinder are greater than the downward force on the top of the cylinder. Their difference is the buoyant

force $\mathbf{F}_{\mathrm{B}}$. (Horizontal forces cancel.)

Just how great is this buoyant force? To answer this question, think about what happens when a submerged object is removed from a fluid, as in [link].



(a) An object submerged in a fluid experiences a buoyant force F_{B}. If F_{B} is greater than the weight of the object, the object will rise. If F_{B} is less than the weight of the object, the object will sink. (b) If the object is removed, it is replaced by fluid having weight w_{fl}. Since this weight is supported by surrounding fluid, the buoyant force must equal the weight of the fluid displaced. That is, $F_{\mathrm{B}} = w_{\mathrm{fl}}$,a statement of Archimedes' principle.

The space it occupied is filled by fluid having a weight w_{fl}. This weight is supported by the surrounding fluid, and so the buoyant force must equal w_{fl} , the weight of the fluid displaced by the object. It is a tribute to the genius

of the Greek mathematician and inventor Archimedes (ca. 287–212 B.C.) that he stated this principle long before concepts of force were well established. Stated in words, **Archimedes' principle** is as follows: The buoyant force on an object equals the weight of the fluid it displaces. In equation form, Archimedes' principle is

Equation:

$$F_{\mathrm{B}} = w_{\mathrm{fl}},$$

where F_{B} is the buoyant force and w_{fl} is the weight of the fluid displaced by the object. Archimedes' principle is valid in general, for any object in any fluid, whether partially or totally submerged.

Note:
Archimedes' Principle
According to this principle the buoyant force on an object equals the weight of the fluid it displaces. In equation form, Archimedes' principle is

Equation:

$$F_{\mathrm{B}} = w_{\mathrm{fl}},$$

where F_{B} is the buoyant force and w_{fl} is the weight of the fluid displaced by the object.

Humm … High-tech body swimsuits were introduced in 2008 in preparation for the Beijing Olympics. One concern (and international rule) was that these suits should not provide any buoyancy advantage. How do you think that this rule could be verified?

Note:
Making Connections: Take-Home Investigation

The density of aluminum foil is 2.7 times the density of water. Take a piece of foil, roll it up into a ball and drop it into water. Does it sink? Why or why not? Can you make it sink?

Floating and Sinking

Drop a lump of clay in water. It will sink. Then mold the lump of clay into the shape of a boat, and it will float. Because of its shape, the boat displaces more water than the lump and experiences a greater buoyant force. The same is true of steel ships.

Example:
Calculating buoyant force: dependency on shape
(a) Calculate the buoyant force on 10,000 metric tons $(1.00 \times 10^7 \text{ kg})$ of solid steel completely submerged in water, and compare this with the steel's weight. (b) What is the maximum buoyant force that water could exert on this same steel if it were shaped into a boat that could displace $1.00 \times 10^5 \text{ m}^3$ of water?
Strategy for (a)
To find the buoyant force, we must find the weight of water displaced. We can do this by using the densities of water and steel given in [link]. We note that, since the steel is completely submerged, its volume and the water's volume are the same. Once we know the volume of water, we can find its mass and weight.
Solution for (a)
First, we use the definition of density $\rho = \frac{m}{V}$ to find the steel's volume, and then we substitute values for mass and density. This gives
Equation:

$$V_{\text{st}} = \frac{m_{\text{st}}}{\rho_{\text{st}}} = \frac{1.00 \times 10^7 \text{ kg}}{7.8 \times 10^3 \text{ kg/m}^3} = 1.28 \times 10^3 \text{ m}^3.$$

Because the steel is completely submerged, this is also the volume of water displaced, V_{w}. We can now find the mass of water displaced from the relationship between its volume and density, both of which are known. This gives

Equation:

$$\begin{aligned} m_{\text{w}} &= \rho_{\text{w}} V_{\text{w}} = (1.000 \times 10^3 \text{ kg/m}^3)(1.28 \times 10^3 \text{ m}^3) \\ &= 1.28 \times 10^6 \text{ kg}. \end{aligned}$$

By Archimedes' principle, the weight of water displaced is $m_{\text{w}}g$, so the buoyant force is

Equation:

$$\begin{aligned} F_{\text{B}} &= w_{\text{w}} = m_{\text{w}}g = \left(1.28 \times 10^6 \text{ kg}\right)\left(9.80 \text{ m/s}^2\right) \\ &= 1.3 \times 10^7 \text{ N}. \end{aligned}$$

The steel's weight is $m_{\text{w}}g = 9.80 \times 10^7$ N, which is much greater than the buoyant force, so the steel will remain submerged. Note that the buoyant force is rounded to two digits because the density of steel is given to only two digits.

Strategy for (b)

Here we are given the maximum volume of water the steel boat can displace. The buoyant force is the weight of this volume of water.

Solution for (b)

The mass of water displaced is found from its relationship to density and volume, both of which are known. That is,

Equation:

$$\begin{aligned} m_{\text{w}} &= \rho_{\text{w}} V_{\text{w}} = \left(1.000 \times 10^3 \text{ kg/m}^3\right)\left(1.00 \times 10^5 \text{ m}^3\right) \\ &= 1.00 \times 10^8 \text{ kg}. \end{aligned}$$

The maximum buoyant force is the weight of this much water, or

Equation:

$$\begin{aligned} F_{\text{B}} &= w_{\text{w}} = m_{\text{w}} g = \left(1.00 \times 10^8 \text{ kg}\right)\left(9.80 \text{ m/s}^2\right) \\ &= 9.80 \times 10^8 \text{ N}. \end{aligned}$$

Discussion
The maximum buoyant force is ten times the weight of the steel, meaning the ship can carry a load nine times its own weight without sinking.

Note:
Making Connections: Take-Home Investigation
A piece of household aluminum foil is 0.016 mm thick. Use a piece of foil that measures 10 cm by 15 cm. (a) What is the mass of this amount of foil? (b) If the foil is folded to give it four sides, and paper clips or washers are added to this “boat,” what shape of the boat would allow it to hold the most “cargo” when placed in water? Test your prediction.

Density and Archimedes’ Principle

Density plays a crucial role in Archimedes’ principle. The average density of an object is what ultimately determines whether it floats. If its average density is less than that of the surrounding fluid, it will float. This is because the fluid, having a higher density, contains more mass and hence more weight in the same volume. The buoyant force, which equals the weight of the fluid displaced, is thus greater than the weight of the object. Likewise, an object denser than the fluid will sink.

The extent to which a floating object is submerged depends on how the object’s density is related to that of the fluid. In [link], for example, the unloaded ship has a lower density and less of it is submerged compared with the same ship loaded. We can derive a quantitative expression for the fraction submerged by considering density. The fraction submerged is the ratio of the volume submerged to the volume of the object, or
Equation:

$$\text{fraction submerged} = \frac{V_{\text{sub}}}{V_{\text{obj}}} = \frac{V_{\text{fl}}}{V_{\text{obj}}}.$$

The volume submerged equals the volume of fluid displaced, which we call V_{fl}. Now we can obtain the relationship between the densities by substituting $\rho = \frac{m}{V}$ into the expression. This gives

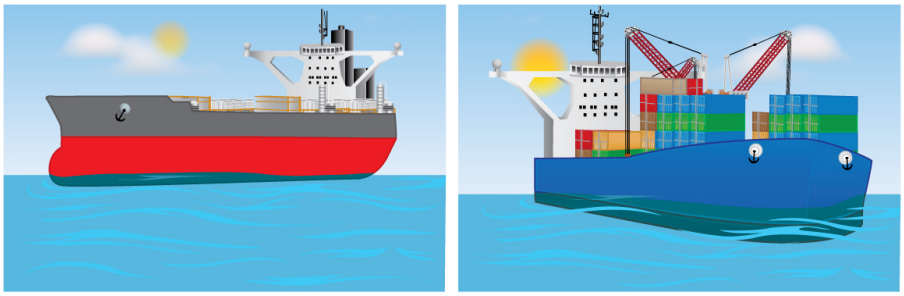
Equation:

$$\frac{V_{\text{fl}}}{V_{\text{obj}}} = \frac{m_{\text{fl}}/\rho_{\text{fl}}}{m_{\text{obj}}/\rho_{\text{obj}}},$$

where ρ_{obj} is the average density of the object and ρ_{fl} is the density of the fluid. Since the object floats, its mass and that of the displaced fluid are equal, and so they cancel from the equation, leaving

Equation:

$$\text{fraction submerged} = \frac{\rho_{\text{obj}}}{\rho_{\text{fl}}}.$$



An unloaded ship (a) floats higher in the water than a loaded ship (b).

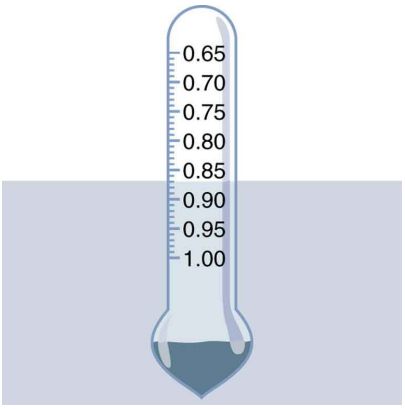
We use this last relationship to measure densities. This is done by measuring the fraction of a floating object that is submerged—for example, with a hydrometer. It is useful to define the ratio of the density of an object to a fluid (usually water) as **specific gravity**:

Equation:

$$\text{specific gravity} = \frac{\rho}{\rho_{\text{w}}},$$

where ρ is the average density of the object or substance and ρ_{w} is the density of water at 4.00°C. Specific gravity is dimensionless, independent of whatever units are used for ρ. If an object floats, its specific gravity is less than one. If it sinks, its specific gravity is greater than one. Moreover, the fraction of a floating object that is submerged equals its specific gravity. If an object's specific gravity is exactly 1, then it will remain suspended in the fluid, neither sinking nor floating. Scuba divers try to obtain this state so that they can hover in the water. We measure the specific gravity of fluids, such as battery acid, radiator fluid, and urine, as an indicator of their condition. One device for measuring specific gravity is shown in [link].

Note:
Specific Gravity
Specific gravity is the ratio of the density of an object to a fluid (usually water).



This hydrometer is floating in a fluid of specific gravity 0.87. The glass hydrometer is filled with air and weighted with lead at the bottom. It floats highest in the densest fluids and has been calibrated and labeled so that specific gravity can be read from it directly.

Example:
Calculating Average Density: Floating Woman
Suppose a 60.0-kg woman floats in freshwater with 97.0% of her volume submerged when her lungs are full of air. What is her average density?
Strategy
We can find the woman's density by solving the equation
Equation:

$$\text{fraction submerged} = \frac{\rho_{\text{obj}}}{\rho_{\text{fl}}}$$

for the density of the object. This yields

Equation:

$$\rho_{\text{obj}} = \rho_{\text{person}} = (\text{fraction submerged}) \cdot \rho_{\text{fl}}.$$

We know both the fraction submerged and the density of water, and so we can calculate the woman’s density.

Solution

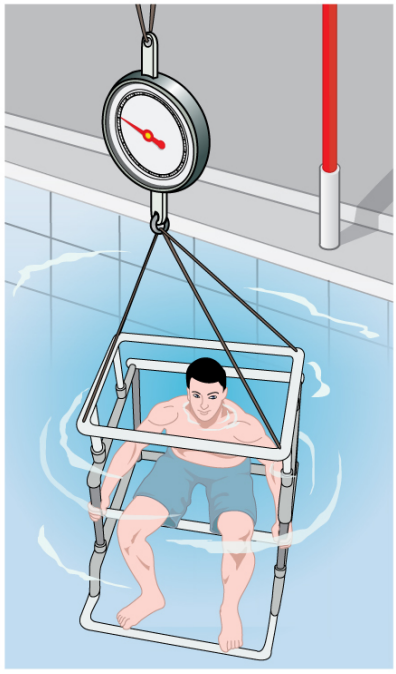
Entering the known values into the expression for her density, we obtain

Equation:

$$\rho_{\text{person}} = 0.970 \cdot \left(10^3 \frac{\text{kg}}{\text{m}^3}\right) = 970 \frac{\text{kg}}{\text{m}^3}.$$

Discussion

Her density is less than the fluid density. We expect this because she floats. Body density is one indicator of a person’s percent body fat, of interest in medical diagnostics and athletic training. (See [link].)



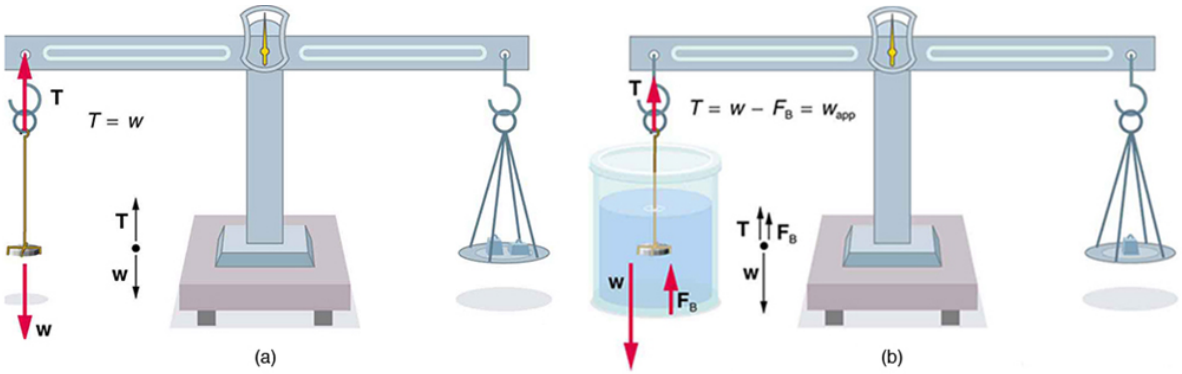
Subject in a “fat tank,” where he is weighed while completely submerged as part of a body density determination. The subject must completely empty his lungs and hold a metal weight in order to sink. Corrections are made for the residual air in his lungs (measured separately) and the metal weight. His corrected submerged weight, his weight in air,

and pinch tests of strategic fatty areas are used to calculate his percent body fat.

There are many obvious examples of lower-density objects or substances floating in higher-density fluids—oil on water, a hot-air balloon, a bit of cork in wine, an iceberg, and hot wax in a “lava lamp,” to name a few. Less obvious examples include lava rising in a volcano and mountain ranges floating on the higher-density crust and mantle beneath them. Even seemingly solid Earth has fluid characteristics.

More Density Measurements

One of the most common techniques for determining density is shown in [link].



(a) A coin is weighed in air. (b) The apparent weight of the coin is determined while it is completely submerged in a fluid of known density. These two measurements are used to calculate the density of the coin.

An object, here a coin, is weighed in air and then weighed again while submerged in a liquid. The density of the coin, an indication of its authenticity, can be calculated if the fluid density is known. This same technique can also be used to determine the density of the fluid if the density of the coin is known. All of these calculations are based on Archimedes' principle.

Archimedes' principle states that the buoyant force on the object equals the weight of the fluid displaced. This, in turn, means that the object *appears* to weigh less when submerged; we call this measurement the object's *apparent weight*. The object suffers an *apparent weight loss* equal to the weight of the fluid displaced. Alternatively, on balances that measure mass, the object suffers an *apparent mass loss* equal to the mass of fluid displaced. That is

Equation:

$$\text{apparent weight loss} = \text{weight of fluid displaced}$$

or

Equation:

$$\text{apparent mass loss} = \text{mass of fluid displaced.}$$

The next example illustrates the use of this technique.

Example:

Calculating Density: Is the Coin Authentic?

The mass of an ancient Greek coin is determined in air to be 8.630 g. When the coin is submerged in water as shown in [link], its apparent mass is 7.800 g. Calculate its density, given that water has a density of $1.000\ \text{g/cm}^3$ and that effects caused by the wire suspending the coin are negligible.

Strategy

To calculate the coin's density, we need its mass (which is given) and its volume. The volume of the coin equals the volume of water displaced. The volume of water displaced V_{w} can be found by solving the equation for density $\rho = \frac{m}{V}$ for V.

Solution

The volume of water is $V_{\text{w}} = \frac{m_{\text{w}}}{\rho_{\text{w}}}$ where m_{w} is the mass of water displaced. As noted, the mass of the water displaced equals the apparent mass loss, which is $m_{\text{w}} = 8.630\ \text{g} - 7.800\ \text{g} = 0.830\ \text{g}$. Thus the volume of water is $V_{\text{w}} = \frac{0.830\ \text{g}}{1.000\ \text{g/cm}^3} = 0.830\ \text{cm}^3$. This is also the volume of the coin, since it is completely submerged. We can now find the density of the coin using the definition of density:

Equation:

$$\rho_{\text{c}} = \frac{m_{\text{c}}}{V_{\text{c}}} = \frac{8.630\ \text{g}}{0.830\ \text{cm}^3} = 10.4\ \text{g/cm}^3.$$

Discussion

You can see from [link] that this density is very close to that of pure silver, appropriate for this type of ancient coin. Most modern counterfeits are not pure silver.

This brings us back to Archimedes' principle and how it came into being. As the story goes, the king of Syracuse gave Archimedes the task of determining whether the royal crown maker was supplying a crown of pure gold. The purity of gold is difficult to determine by color (it can be diluted with other metals and still look as yellow as pure gold), and other analytical techniques had not yet been conceived. Even ancient peoples, however, realized that the density of gold was greater than that of any other then-known substance. Archimedes purportedly agonized over his task and had his inspiration one day while at the public baths, pondering the support the water gave his body. He came up with his now-famous principle, saw how to apply it to determine density, and ran naked down the streets of Syracuse crying "Eureka!" (Greek for "I have found it"). Similar behavior can be observed in contemporary physicists from time to time!

Note:
PhET Explorations: Buoyancy
When will objects float and when will they sink? Learn how buoyancy works with blocks. Arrows show the applied forces, and you can modify the properties of the blocks and the fluid.
https://phet.colorado.edu/sims/density-and-buoyancy/buoyancy_en.html

Section Summary

- Buoyant force is the net upward force on any object in any fluid. If the buoyant force is greater than the object's weight, the object will rise to the surface and float. If the buoyant force is less than the object's weight, the object will sink. If the buoyant force equals the object's weight, the object will remain suspended at that depth. The buoyant force is always present whether the object floats, sinks, or is suspended in a fluid.
- Archimedes' principle states that the buoyant force on an object equals the weight of the fluid it displaces.
- Specific gravity is the ratio of the density of an object to a fluid (usually water).

Conceptual Questions

Exercise:

Problem:

More force is required to pull the plug in a full bathtub than when it is empty. Does this contradict Archimedes' principle? Explain your answer.

Exercise:

Problem:

Do fluids exert buoyant forces in a "weightless" environment, such as in the space shuttle? Explain your answer.

Exercise:

Problem:

Will the same ship float higher in salt water than in freshwater? Explain your answer.

Exercise:

Problem:

Marbles dropped into a partially filled bathtub sink to the bottom. Part of their weight is supported by buoyant force, yet the downward force on the bottom of the tub increases by exactly the weight of the marbles. Explain why.

Problem Exercises

Exercise:

Problem:

What fraction of ice is submerged when it floats in freshwater, given the density of water at 0°C is very close to $1000\ \mathrm{kg/m^3}$?

Solution:

91.7%

Exercise:

Problem:

Logs sometimes float vertically in a lake because one end has become water-logged and denser than the other. What is the average density of a uniform-diameter log that floats with 20.0% of its length above water?

Exercise:

Problem:

Find the density of a fluid in which a hydrometer having a density of $0.750\ \mathrm{g/mL}$ floats with 92.0% of its volume submerged.

Solution:

$815\ \mathrm{kg/m^3}$

Exercise:

Problem:

If your body has a density of $995\ \mathrm{kg/m^3}$, what fraction of you will be submerged when floating gently in: (a) freshwater? (b) salt water, which has a density of $1027\ \mathrm{kg/m^3}$?

Exercise:

Problem:

Bird bones have air pockets in them to reduce their weight—this also gives them an average density significantly less than that of the bones of other animals. Suppose an ornithologist weighs a bird bone in air and in water and finds its mass is $45.0\ \mathrm{g}$ and its apparent mass when submerged is $3.60\ \mathrm{g}$ (the bone is watertight). (a) What mass of water is displaced? (b) What is the volume of the bone? (c) What is its average density?

Solution:

(a) 41.4 g

(b) $41.4\ \mathrm{cm^3}$

(c) $1.09\ \mathrm{g/cm^3}$

Exercise:

Problem:

A rock with a mass of 540 g in air is found to have an apparent mass of 342 g when submerged in water. (a) What mass of water is displaced? (b) What is the volume of the rock? (c) What is its average density? Is this consistent with the value for granite?

Exercise:

Problem:

Archimedes' principle can be used to calculate the density of a fluid as well as that of a solid. Suppose a chunk of iron with a mass of 390.0 g in air is found to have an apparent mass of 350.5 g when completely submerged in an unknown liquid. (a) What mass of fluid does the iron displace? (b) What is the volume of iron, using its density as given in [link] (c) Calculate the fluid's density and identify it.

Solution:

(a) 39.5 g

(b) $50\ \text{cm}^3$

(c) $0.79\ \text{g/cm}^3$

It is ethyl alcohol.

Exercise:

Problem:

In an immersion measurement of a woman's density, she is found to have a mass of 62.0 kg in air and an apparent mass of 0.0850 kg when completely submerged with lungs empty. (a) What mass of water does she displace? (b) What is her volume? (c) Calculate her density. (d) If her lung capacity is 1.75 L, is she able to float without treading water with her lungs filled with air?

Exercise:

Problem:

Some fish have a density slightly less than that of water and must exert a force (swim) to stay submerged. What force must an 85.0-kg grouper exert to stay submerged in salt water if its body density is $1015\ \mathrm{kg/m^3}$?

Solution:

8.21 N

Exercise:

Problem:

(a) Calculate the buoyant force on a 2.00-L helium balloon. (b) Given the mass of the rubber in the balloon is 1.50 g, what is the net vertical force on the balloon if it is let go? You can neglect the volume of the rubber.

Exercise:

Problem:

(a) What is the density of a woman who floats in freshwater with 4.00% of her volume above the surface? This could be measured by placing her in a tank with marks on the side to measure how much water she displaces when floating and when held under water (briefly). (b) What percent of her volume is above the surface when she floats in seawater?

Solution:

(a) $960\ \mathrm{kg/m^3}$

(b) 6.34%

She indeed floats more in seawater.

Exercise:

Problem:

A certain man has a mass of 80 kg and a density of $955\ \mathrm{kg/m^3}$ (excluding the air in his lungs). (a) Calculate his volume. (b) Find the buoyant force air exerts on him. (c) What is the ratio of the buoyant force to his weight?

Exercise:

Problem:

A simple compass can be made by placing a small bar magnet on a cork floating in water. (a) What fraction of a plain cork will be submerged when floating in water? (b) If the cork has a mass of 10.0 g and a 20.0-g magnet is placed on it, what fraction of the cork will be submerged? (c) Will the bar magnet and cork float in ethyl alcohol?

Solution:

(a) 0.24

(b) 0.68

(c) Yes, the cork will float because
$\rho_{\text{obj}} < \rho_{\text{ethyl alcohol}} (0.678\ \mathrm{g/cm^3} < 0.79\ \mathrm{g/cm^3})$

Exercise:

Problem:

What fraction of an iron anchor’s weight will be supported by buoyant force when submerged in saltwater?

Exercise:

Problem:

Scurrilous con artists have been known to represent gold-plated tungsten ingots as pure gold and sell them to the greedy at prices much below gold value but deservedly far above the cost of tungsten. With what accuracy must you be able to measure the mass of such an ingot in and out of water to tell that it is almost pure tungsten rather than pure gold?

Solution:

The difference is 0.006%.

Exercise:

Problem:

A twin-sized air mattress used for camping has dimensions of 100 cm by 200 cm by 15 cm when blown up. The weight of the mattress is 2 kg. How heavy a person could the air mattress hold if it is placed in freshwater?

Exercise:

Problem:

Referring to [link], prove that the buoyant force on the cylinder is equal to the weight of the fluid displaced (Archimedes' principle). You may assume that the buoyant force is $F_2 - F_1$ and that the ends of the cylinder have equal areas A. Note that the volume of the cylinder (and that of the fluid it displaces) equals $(h_2 - h_1)A$.

Solution:

$$F_{\text{net}} = F_2 - F_1 = P_2A - P_1A = (P_2 - P_1)A$$

$$= (h_2\rho_{\text{fl}}g - h_1\rho_{\text{fl}}g)A$$

$$= (h_2 - h_1)\rho_{\text{fl}}\text{gA}$$

where ρ_{fl} = density of fluid. Therefore,

$$F_{\text{net}} = (h_2 - h_1)A\rho_{\text{fl}}g = V_{\text{fl}}\rho_{\text{fl}}g = m_{\text{fl}}g = w_{\text{fl}}$$

where is w_{fl} the weight of the fluid displaced.

Exercise:

Problem:

(a) A 75.0-kg man floats in freshwater with 3.00% of his volume above water when his lungs are empty, and 5.00% of his volume above water when his lungs are full. Calculate the volume of air he inhales—called his lung capacity—in liters. (b) Does this lung volume seem reasonable?

Glossary

Archimedes' principle
: the buoyant force on an object equals the weight of the fluid it displaces

buoyant force
: the net upward force on any object in any fluid

specific gravity
: the ratio of the density of an object to a fluid (usually water)

Cohesion and Adhesion in Liquids: Surface Tension and Capillary Action

- Understand cohesive and adhesive forces.
- Define surface tension.
- Understand capillary action.

Cohesion and Adhesion in Liquids

Children blow soap bubbles and play in the spray of a sprinkler on a hot summer day. (See [link].) An underwater spider keeps his air supply in a shiny bubble he carries wrapped around him. A technician draws blood into a small-diameter tube just by touching it to a drop on a pricked finger. A premature infant struggles to inflate her lungs. What is the common thread? All these activities are dominated by the attractive forces between atoms and molecules in liquids—both within a liquid and between the liquid and its surroundings.

Attractive forces between molecules of the same type are called **cohesive forces**. Liquids can, for example, be held in open containers because cohesive forces hold the molecules together. Attractive forces between molecules of different types are called **adhesive forces**. Such forces cause liquid drops to cling to window panes, for example. In this section we examine effects directly attributable to cohesive and adhesive forces in liquids.

Note:
Cohesive Forces
Attractive forces between molecules of the same type are called cohesive forces.

Note:
Adhesive Forces

Attractive forces between molecules of different types are called adhesive forces.



The soap bubbles in this photograph are caused by cohesive forces among molecules in liquids. (credit: Steven Depolo, Flickr)

Surface Tension

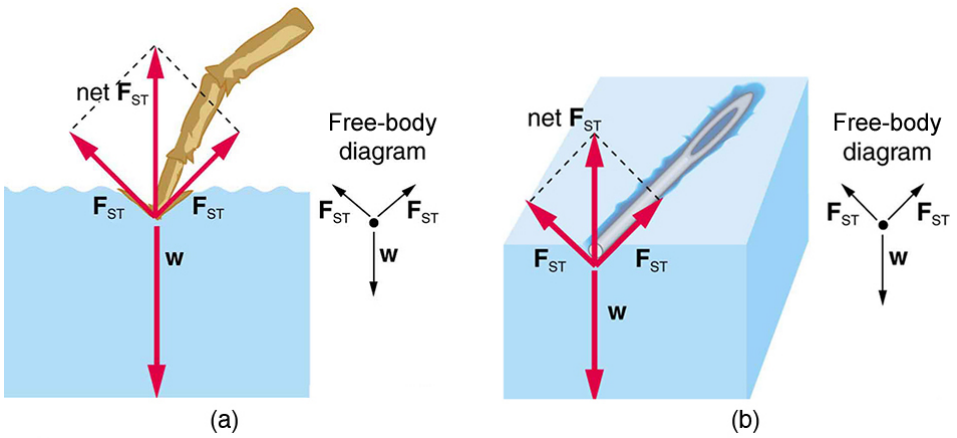
Cohesive forces between molecules cause the surface of a liquid to contract to the smallest possible surface area. This general effect is called **surface tension**. Molecules on the surface are pulled inward by cohesive forces, reducing the surface area. Molecules inside the liquid experience zero net force, since they have neighbors on all sides.

Note:
Surface Tension

Cohesive forces between molecules cause the surface of a liquid to contract to the smallest possible surface area. This general effect is called surface tension.

Note:
Making Connections: Surface Tension
Forces between atoms and molecules underlie the macroscopic effect called surface tension. These attractive forces pull the molecules closer together and tend to minimize the surface area. This is another example of a submicroscopic explanation for a macroscopic phenomenon.

The model of a liquid surface acting like a stretched elastic sheet can effectively explain surface tension effects. For example, some insects can walk on water (as opposed to floating in it) as we would walk on a trampoline—they dent the surface as shown in [link](a). [link](b) shows another example, where a needle rests on a water surface. The iron needle cannot, and does not, float, because its density is greater than that of water. Rather, its weight is supported by forces in the stretched surface that try to make the surface smaller or flatter. If the needle were placed point down on the surface, its weight acting on a smaller area would break the surface, and it would sink.



Surface tension supporting the weight of an insect and an iron needle, both of which rest on the surface without penetrating it. They are not floating; rather, they are supported by the surface of the liquid. (a) An insect leg dents the water surface. F_{ST} is a restoring force (surface tension) parallel to the surface. (b) An iron needle similarly dents a water surface until the restoring force (surface tension) grows to equal its weight.

Surface tension is proportional to the strength of the cohesive force, which varies with the type of liquid. Surface tension γ is defined to be the force F per unit length L exerted by a stretched liquid membrane:

Equation:

$$\gamma = \frac{F}{L}.$$

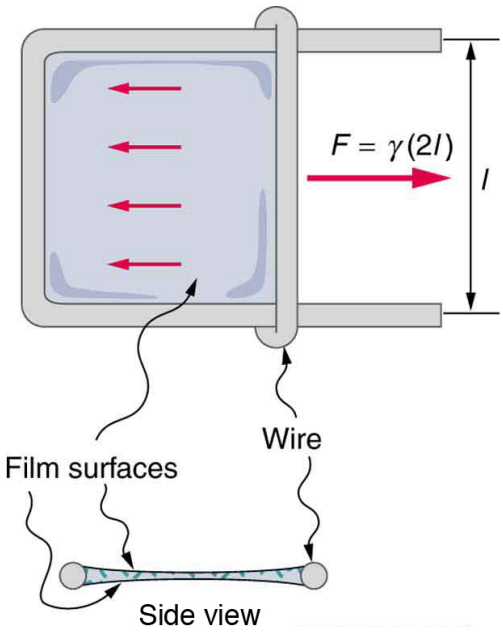
[link] lists values of γ for some liquids. For the insect of [link](a), its weight w is supported by the upward components of the surface tension force: $w = \gamma L \sin\theta$, where L is the circumference of the insect's foot in contact with the water. [link] shows one way to measure surface tension. The liquid film exerts a force on the movable wire in an attempt to reduce its surface area. The magnitude of this force depends on the surface tension of the liquid and can be measured accurately.

Surface tension is the reason why liquids form bubbles and droplets. The inward surface tension force causes bubbles to be approximately spherical and raises the pressure of the gas trapped inside relative to atmospheric pressure outside. It can be shown that the gauge pressure P inside a spherical bubble is given by

Equation:

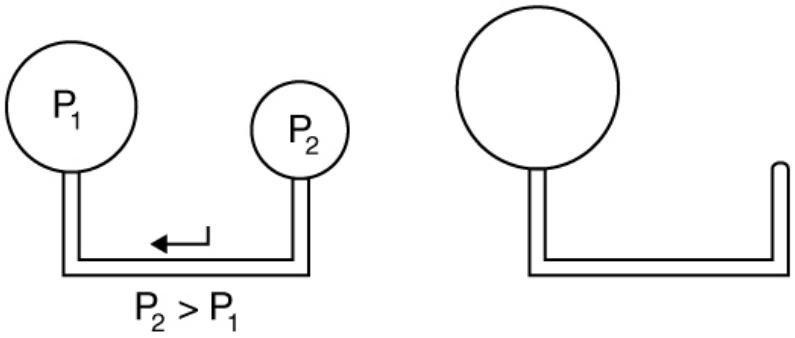
$$P = \frac{4\gamma}{r},$$

where r is the radius of the bubble. Thus the pressure inside a bubble is greatest when the bubble is the smallest. Another bit of evidence for this is illustrated in [link]. When air is allowed to flow between two balloons of unequal size, the smaller balloon tends to collapse, filling the larger balloon.



Sliding wire device used for measuring surface tension; the device exerts a force to reduce the film's surface area. The force needed to hold the wire in place is $F = \gamma L = \gamma(2l)$, since there are *two* liquid surfaces attached to the wire. This force remains nearly constant as the

film is stretched, until the film approaches its breaking point.



With the valve closed, two balloons of different sizes are attached to each end of a tube. Upon opening the valve, the smaller balloon decreases in size with the air moving to fill the larger balloon. The pressure in a spherical balloon is inversely proportional to its radius, so that the smaller balloon has a greater internal pressure than the larger balloon, resulting in this flow.

Liquid	**Surface tension γ(N/m)**
Water at 0°C	0.0756

Liquid	Surface tension γ(N/m)
Water at 20ºC	0.0728
Water at 100ºC	0.0589
Soapy water (typical)	0.0370
Ethyl alcohol	0.0223
Glycerin	0.0631
Mercury	0.465
Olive oil	0.032
Tissue fluids (typical)	0.050
Blood, whole at 37ºC	0.058
Blood plasma at 37ºC	0.073
Gold at 1070ºC	1.000
Oxygen at −193ºC	0.0157
Helium at −269ºC	0.00012

Surface Tension of Some Liquids[footnote]
At 20ºC unless otherwise stated.

Example:
Surface Tension: Pressure Inside a Bubble
Calculate the gauge pressure inside a soap bubble 2.00×10^{-4} m in radius using the surface tension for soapy water in [link]. Convert this pressure to

mm Hg.

Strategy

The radius is given and the surface tension can be found in [link], and so P can be found directly from the equation $P = \frac{4\gamma}{r}$.

Solution

Substituting r and γ into the equation $P = \frac{4\gamma}{r}$, we obtain

Equation:

$$P = \frac{4\gamma}{r} = \frac{4(0.037\ \text{N/m})}{2.00 \times 10^{-4}\ \text{m}} = 740\ \text{N/m}^2 = 740\ \text{Pa}.$$

We use a conversion factor to get this into units of mm Hg:

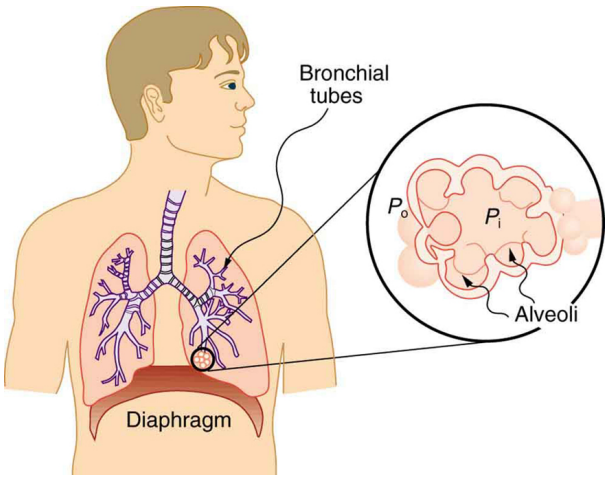
Equation:

$$P = \left(740\ \text{N/m}^2\right)\frac{1.00\ \text{mm Hg}}{133\ \text{N/m}^2} = 5.56\ \text{mm Hg}.$$

Discussion

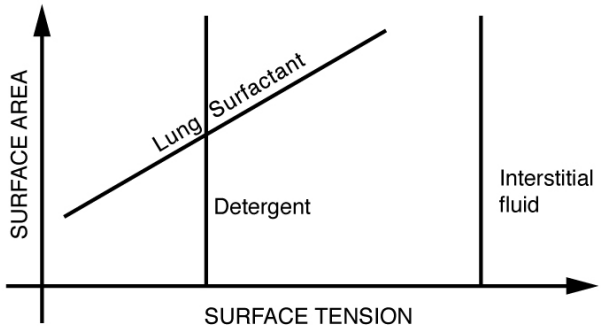
Note that if a hole were to be made in the bubble, the air would be forced out, the bubble would decrease in radius, and the pressure inside would *increase* to atmospheric pressure (760 mm Hg).

Our lungs contain hundreds of millions of mucus-lined sacs called *alveoli*, which are very similar in size, and about 0.1 mm in diameter. (See [link].) You can exhale without muscle action by allowing surface tension to contract these sacs. Medical patients whose breathing is aided by a positive pressure respirator have air blown into the lungs, but are generally allowed to exhale on their own. Even if there is paralysis, surface tension in the alveoli will expel air from the lungs. Since pressure increases as the radii of the alveoli decrease, an occasional deep cleansing breath is needed to fully reinflate the alveoli. Respirators are programmed to do this and we find it natural, as do our companion dogs and cats, to take a cleansing breath before settling into a nap.



Bronchial tubes in the lungs branch into ever-smaller structures, finally ending in alveoli. The alveoli act like tiny bubbles. The surface tension of their mucous lining aids in exhalation and can prevent inhalation if too great.

The tension in the walls of the alveoli results from the membrane tissue and a liquid on the walls of the alveoli containing a long lipoprotein that acts as a surfactant (a surface-tension reducing substance). The need for the surfactant results from the tendency of small alveoli to collapse and the air to fill into the larger alveoli making them even larger (as demonstrated in [link]). During inhalation, the lipoprotein molecules are pulled apart and the wall tension increases as the radius increases (increased surface tension). During exhalation, the molecules slide back together and the surface tension decreases, helping to prevent a collapse of the alveoli. The surfactant therefore serves to change the wall tension so that small alveoli don’t collapse and large alveoli are prevented from expanding too much. This tension change is a unique property of these surfactants, and is not shared by detergents (which simply lower surface tension). (See [link].)



Surface tension as a function of surface area. The surface tension for lung surfactant decreases with decreasing area. This ensures that small alveoli don’t collapse and large alveoli are not able to over expand.

If water gets into the lungs, the surface tension is too great and you cannot inhale. This is a severe problem in resuscitating drowning victims. A similar problem occurs in newborn infants who are born without this surfactant—their lungs are very difficult to inflate. This condition is known as *hyaline membrane disease* and is a leading cause of death for infants, particularly in premature births. Some success has been achieved in treating hyaline membrane disease by spraying a surfactant into the infant’s breathing passages. Emphysema produces the opposite problem with alveoli. Alveolar walls of emphysema victims deteriorate, and the sacs combine to form larger sacs. Because pressure produced by surface tension decreases with increasing radius, these larger sacs produce smaller pressure, reducing the ability of emphysema victims to exhale. A common test for emphysema is to measure the pressure and volume of air that can be exhaled.

Note:
Making Connections: Take-Home Investigation

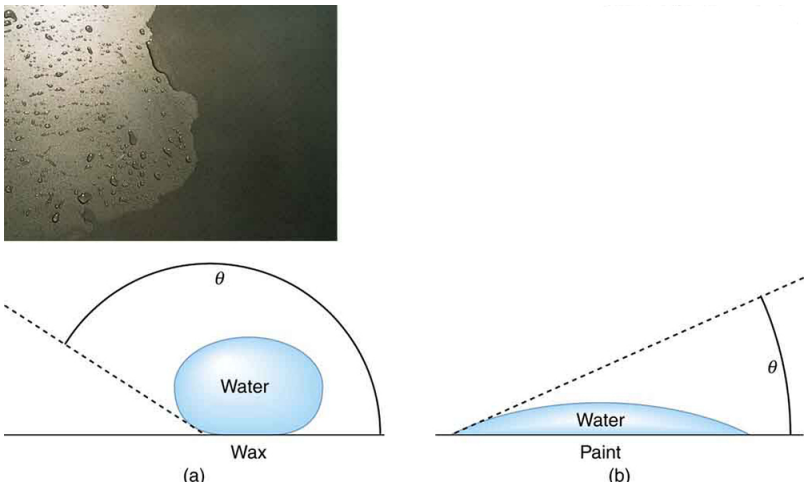
(1) Try floating a sewing needle on water. In order for this activity to work, the needle needs to be very clean as even the oil from your fingers can be sufficient to affect the surface properties of the needle. (2) Place the bristles of a paint brush into water. Pull the brush out and notice that for a short while, the bristles will stick together. The surface tension of the water surrounding the bristles is sufficient to hold the bristles together. As the bristles dry out, the surface tension effect dissipates. (3) Place a loop of thread on the surface of still water in such a way that all of the thread is in contact with the water. Note the shape of the loop. Now place a drop of detergent into the middle of the loop. What happens to the shape of the loop? Why? (4) Sprinkle pepper onto the surface of water. Add a drop of detergent. What happens? Why? (5) Float two matches parallel to each other and add a drop of detergent between them. What happens? Note: For each new experiment, the water needs to be replaced and the bowl washed to free it of any residual detergent.

Adhesion and Capillary Action

Why is it that water beads up on a waxed car but does not on bare paint? The answer is that the adhesive forces between water and wax are much smaller than those between water and paint. Competition between the forces of adhesion and cohesion are important in the macroscopic behavior of liquids. An important factor in studying the roles of these two forces is the angle θ between the tangent to the liquid surface and the surface. (See [link].) The **contact angle** θ is directly related to the relative strength of the cohesive and adhesive forces. The larger the strength of the cohesive force relative to the adhesive force, the larger θ is, and the more the liquid tends to form a droplet. The smaller θ is, the smaller the relative strength, so that the adhesive force is able to flatten the drop. [link] lists contact angles for several combinations of liquids and solids.

Note:
Contact Angle

The angle θ between the tangent to the liquid surface and the surface is called the contact angle.



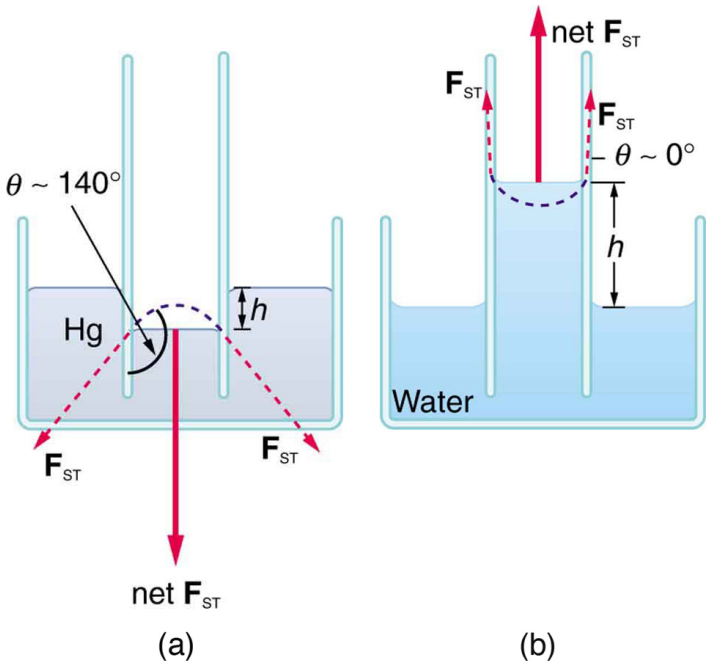
In the photograph, water beads on the waxed car paint and flattens on the unwaxed paint. (a) Water forms beads on the waxed surface because the cohesive forces responsible for surface tension are larger than the adhesive forces, which tend to flatten the drop. (b) Water beads on bare paint are flattened considerably because the adhesive forces between water and paint are strong, overcoming surface tension. The contact angle θ is directly related to the relative strengths of the cohesive and adhesive forces. The larger θ is, the larger the ratio of cohesive to adhesive forces. (credit: P. P. Urone)

One important phenomenon related to the relative strength of cohesive and adhesive forces is **capillary action**—the tendency of a fluid to be raised or

suppressed in a narrow tube, or *capillary tube*. This action causes blood to be drawn into a small-diameter tube when the tube touches a drop.

Note:
Capillary Action
The tendency of a fluid to be raised or suppressed in a narrow tube, or capillary tube, is called capillary action.

If a capillary tube is placed vertically into a liquid, as shown in [link], capillary action will raise or suppress the liquid inside the tube depending on the combination of substances. The actual effect depends on the relative strength of the cohesive and adhesive forces and, thus, the contact angle θ given in the table. If θ is less than 90°, then the fluid will be raised; if θ is greater than 90°, it will be suppressed. Mercury, for example, has a very large surface tension and a large contact angle with glass. When placed in a tube, the surface of a column of mercury curves downward, somewhat like a drop. The curved surface of a fluid in a tube is called a **meniscus**. The tendency of surface tension is always to reduce the surface area. Surface tension thus flattens the curved liquid surface in a capillary tube. This results in a downward force in mercury and an upward force in water, as seen in [link].



(a) Mercury is suppressed in a glass tube because its contact angle is greater than 90º. Surface tension exerts a downward force as it flattens the mercury, suppressing it in the tube. The dashed line shows the shape the mercury surface would have without the flattening effect of surface tension. (b) Water is raised in a glass tube because its contact angle is nearly 0º. Surface tension therefore exerts an upward force when it flattens the surface to reduce its area.

Interface	Contact angle θ
Mercury–glass	140°
Water–glass	0°
Water–paraffin	107°
Water–silver	90°
Organic liquids (most)–glass	0°
Ethyl alcohol–glass	0°
Kerosene–glass	26°

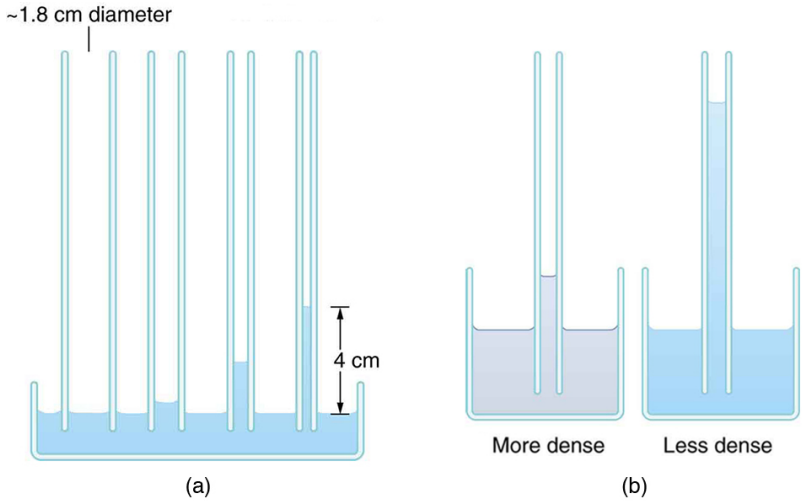
Contact Angles of Some Substances

Capillary action can move liquids horizontally over very large distances, but the height to which it can raise or suppress a liquid in a tube is limited by its weight. It can be shown that this height h is given by

Equation:

$$h = \frac{2\gamma \cos\theta}{\rho g r}.$$

If we look at the different factors in this expression, we might see how it makes good sense. The height is directly proportional to the surface tension γ, which is its direct cause. Furthermore, the height is inversely proportional to tube radius—the smaller the radius r, the higher the fluid can be raised, since a smaller tube holds less mass. The height is also inversely proportional to fluid density ρ, since a larger density means a greater mass in the same volume. (See [link].)



(a) Capillary action depends on the radius of a tube. The smaller the tube, the greater the height reached. The height is negligible for large-radius tubes. (b) A denser fluid in the same tube rises to a smaller height, all other factors being the same.

Example:
Calculating Radius of a Capillary Tube: Capillary Action: Tree Sap

Can capillary action be solely responsible for sap rising in trees? To answer this question, calculate the radius of a capillary tube that would raise sap 100 m to the top of a giant redwood, assuming that sap's density is $1050\ \text{kg/m}^3$, its contact angle is zero, and its surface tension is the same as that of water at $20.0^\text{o}\ \text{C}$.

Strategy

The height to which a liquid will rise as a result of capillary action is given by $h = \frac{2\gamma \cos \theta}{\rho g r}$, and every quantity is known except for r.

Solution

Solving for r and substituting known values produces

Equation:

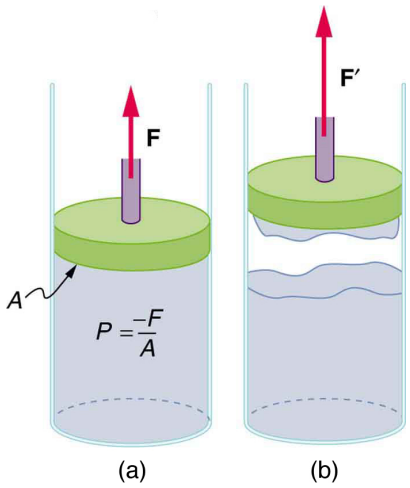
$$\begin{aligned} r &= \frac{2\gamma \cos \theta}{\rho g h} = \frac{2(0.0728\ \text{N/m})\cos(0^\text{o})}{\left(1050\ \text{kg/m}^3\right)\left(9.80\ \text{m/s}^2\right)(100\ \text{m})} \\ &= 1.41 \times 10^{-7}\ \text{m}. \end{aligned}$$

Discussion

This result is unreasonable. Sap in trees moves through the *xylem*, which forms tubes with radii as small as 2.5×10^{-5} m. This value is about 180 times as large as the radius found necessary here to raise sap 100 m. This means that capillary action alone cannot be solely responsible for sap getting to the tops of trees.

How *does* sap get to the tops of tall trees? (Recall that a column of water can only rise to a height of 10 m when there is a vacuum at the top—see [link].) The question has not been completely resolved, but it appears that it is pulled up like a chain held together by cohesive forces. As each molecule of sap enters a leaf and evaporates (a process called transpiration), the entire chain is pulled up a notch. So a negative pressure created by water evaporation must be present to pull the sap up through the xylem vessels. In most situations, *fluids can push but can exert only negligible pull*, because the cohesive forces seem to be too small to hold the molecules tightly together. But in this case, the cohesive force of water molecules provides a very strong pull. [link] shows one device for studying negative pressure.

Some experiments have demonstrated that negative pressures sufficient to pull sap to the tops of the tallest trees *can* be achieved.



(a) When the piston is raised, it stretches the liquid slightly, putting it under tension and creating a negative absolute pressure $P = -F/A$. (b) The liquid eventually separates, giving an experimental limit to negative pressure in this liquid.

Section Summary

- Attractive forces between molecules of the same type are called cohesive forces.
- Attractive forces between molecules of different types are called adhesive forces.
- Cohesive forces between molecules cause the surface of a liquid to contract to the smallest possible surface area. This general effect is called surface tension.
- Capillary action is the tendency of a fluid to be raised or suppressed in a narrow tube, or capillary tube which is due to the relative strength of cohesive and adhesive forces.

Conceptual Questions

Exercise:

Problem:

The density of oil is less than that of water, yet a loaded oil tanker sits lower in the water than an empty one. Why?

Exercise:

Problem:

Is surface tension due to cohesive or adhesive forces, or both?

Exercise:

Problem:

Is capillary action due to cohesive or adhesive forces, or both?

Exercise:

Problem:

Birds such as ducks, geese, and swans have greater densities than water, yet they are able to sit on its surface. Explain this ability, noting that water does not wet their feathers and that they cannot sit on soapy water.

Exercise:

Problem:

Water beads up on an oily sunbather, but not on her neighbor, whose skin is not oiled. Explain in terms of cohesive and adhesive forces.

Exercise:

Problem:

Could capillary action be used to move fluids in a “weightless” environment, such as in an orbiting space probe?

Exercise:

Problem:

What effect does capillary action have on the reading of a manometer with uniform diameter? Explain your answer.

Exercise:

Problem:

Pressure between the inside chest wall and the outside of the lungs normally remains negative. Explain how pressure inside the lungs can become positive (to cause exhalation) without muscle action.

Problems & Exercises

Exercise:

Problem:

What is the pressure inside an alveolus having a radius of 2.50×10^{-4} m if the surface tension of the fluid-lined wall is the same as for soapy water? You may assume the pressure is the same as that created by a spherical bubble.

Solution:

$592\ \mathrm{N/m^2}$

Exercise:

Problem:

(a) The pressure inside an alveolus with a 2.00×10^{-4}-m radius is $1.40 \times 10^{3}\ \mathrm{Pa}$, due to its fluid-lined walls. Assuming the alveolus acts like a spherical bubble, what is the surface tension of the fluid? (b) Identify the likely fluid. (You may need to extrapolate between values in [link].)

Exercise:

Problem:

What is the gauge pressure in millimeters of mercury inside a soap bubble 0.100 m in diameter?

Solution:

$2.23 \times 10^{-2}\ \mathrm{mm\ Hg}$

Exercise:

Problem:

Calculate the force on the slide wire in [link] if it is 3.50 cm long and the fluid is ethyl alcohol.

Exercise:

Problem:

[link](a) shows the effect of tube radius on the height to which capillary action can raise a fluid. (a) Calculate the height h for water in a glass tube with a radius of 0.900 cm—a rather large tube like the one on the left. (b) What is the radius of the glass tube on the right if it raises water to 4.00 cm?

Solution:

(a) 1.65×10^{-3} m

(b) 3.71×10^{-4} m

Exercise:

Problem:

We stated in [link] that a xylem tube is of radius 2.50×10^{-5} m. Verify that such a tube raises sap less than a meter by finding h for it, making the same assumptions that sap's density is $1050\ \text{kg/m}^3$, its contact angle is zero, and its surface tension is the same as that of water at 20.0° C.

Exercise:

Problem:

What fluid is in the device shown in [link] if the force is 3.16×10^{-3} N and the length of the wire is 2.50 cm? Calculate the surface tension γ and find a likely match from [link].

Solution:

$6.32 \times 10^{-2}\ \text{N/m}$

Based on the values in table, the fluid is probably glycerin.

Exercise:

Problem:

If the gauge pressure inside a rubber balloon with a 10.0-cm radius is 1.50 cm of water, what is the effective surface tension of the balloon?

Exercise:

Problem:

Calculate the gauge pressures inside 2.00-cm-radius bubbles of water, alcohol, and soapy water. Which liquid forms the most stable bubbles, neglecting any effects of evaporation?

Solution:

$$\begin{aligned} P_{\text{w}} &= 14.6\ \text{N/m}^2, \\ P_{\text{a}} &= 4.46\ \text{N/m}^2, \\ P_{\text{sw}} &= 7.40\ \text{N/m}^2. \end{aligned}$$

Alcohol forms the most stable bubble, since the absolute pressure inside is closest to atmospheric pressure.

Exercise:

Problem:

Suppose water is raised by capillary action to a height of 5.00 cm in a glass tube. (a) To what height will it be raised in a paraffin tube of the same radius? (b) In a silver tube of the same radius?

Exercise:

Problem:

Calculate the contact angle θ for olive oil if capillary action raises it to a height of 7.07 cm in a glass tube with a radius of 0.100 mm. Is this value consistent with that for most organic liquids?

Solution:

5.1^{o}

This is near the value of $\theta = 0^{\text{o}}$ for most organic liquids.

Exercise:

Problem:

When two soap bubbles touch, the larger is inflated by the smaller until they form a single bubble. (a) What is the gauge pressure inside a soap bubble with a 1.50-cm radius? (b) Inside a 4.00-cm-radius soap bubble? (c) Inside the single bubble they form if no air is lost when they touch?

Exercise:

Problem:

Calculate the ratio of the heights to which water and mercury are raised by capillary action in the same glass tube.

Solution:

-2.78

The ratio is negative because water is raised whereas mercury is lowered.

Exercise:

Problem:

What is the ratio of heights to which ethyl alcohol and water are raised by capillary action in the same glass tube?

Glossary

adhesive forces
: the attractive forces between molecules of different types

capillary action
: the tendency of a fluid to be raised or lowered in a narrow tube

cohesive forces
: the attractive forces between molecules of the same type

contact angle
: the angle θ between the tangent to the liquid surface and the surface

surface tension
: the cohesive forces between molecules which cause the surface of a liquid to contract to the smallest possible surface area

Pressures in the Body

- Explain the concept of pressure the in human body.
- Explain systolic and diastolic blood pressures.
- Describe pressures in the eye, lungs, spinal column, bladder, and skeletal system.

Pressure in the Body

Next to taking a person's temperature and weight, measuring blood pressure is the most common of all medical examinations. Control of high blood pressure is largely responsible for the significant decreases in heart attack and stroke fatalities achieved in the last three decades. The pressures in various parts of the body can be measured and often provide valuable medical indicators. In this section, we consider a few examples together with some of the physics that accompanies them.

[link] lists some of the measured pressures in mm Hg, the units most commonly quoted.

Body system	**Gauge pressure in mm Hg**
Blood pressures in large arteries (resting)	
Maximum (systolic)	100–140
Minimum (diastolic)	60–90
Blood pressure in large veins	4–15
Eye	12–24
Brain and spinal fluid (lying down)	5–12
Bladder	
While filling	0–25
When full	100–150
Chest cavity between lungs and ribs	−8 to −4
Inside lungs	−2 to +3
Digestive tract	

Body system	Gauge pressure in mm Hg
Esophagus	−2
Stomach	0–20
Intestines	10–20
Middle ear	<1

Typical Pressures in Humans

Blood Pressure

Common arterial blood pressure measurements typically produce values of 120 mm Hg and 80 mm Hg, respectively, for systolic and diastolic pressures. Both pressures have health implications. When systolic pressure is chronically high, the risk of stroke and heart attack is increased. If, however, it is too low, fainting is a problem. **Systolic pressure** increases dramatically during exercise to increase blood flow and returns to normal afterward. This change produces no ill effects and, in fact, may be beneficial to the tone of the circulatory system. **Diastolic pressure** can be an indicator of fluid balance. When low, it may indicate that a person is hemorrhaging internally and needs a transfusion. Conversely, high diastolic pressure indicates a ballooning of the blood vessels, which may be due to the transfusion of too much fluid into the circulatory system. High diastolic pressure is also an indication that blood vessels are not dilating properly to pass blood through. This can seriously strain the heart in its attempt to pump blood.

Blood leaves the heart at about 120 mm Hg but its pressure continues to decrease (to almost 0) as it goes from the aorta to smaller arteries to small veins (see [link]). The pressure differences in the circulation system are caused by blood flow through the system as well as the position of the person. For a person standing up, the pressure in the feet will be larger than at the heart due to the weight of the blood $(P = h\rho g)$. If we assume that the distance between the heart and the feet of a person in an upright position is 1.4 m, then the increase in pressure in the feet relative to that in the heart (for a static column of blood) is given by

Equation:

$$\Delta P = \Delta h\rho g = (1.4\text{ m})\left(1050\text{ kg/m}^3\right)\left(9.80\text{ m/s}^2\right) = 1.4 \times 10^4\text{ Pa} = 108\text{ mm Hg}.$$

Note:
Increase in Pressure in the Feet of a Person

Equation:

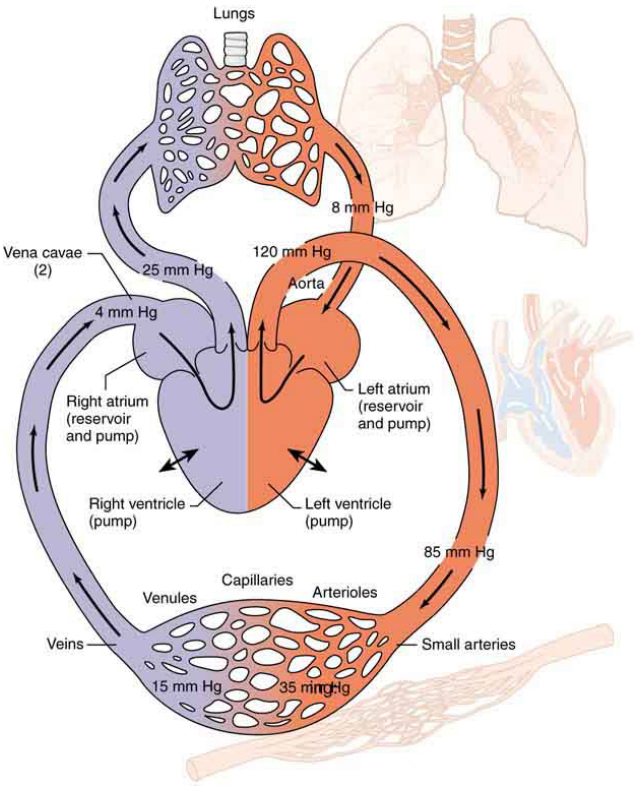
$$\Delta P = \Delta h\rho g = (1.4\text{ m})\left(1050\text{ kg/m}^3\right)\left(9.80\text{ m/s}^2\right) = 1.4 \times 10^4\text{ Pa} = 108\text{ mm Hg}.$$

Standing a long time can lead to an accumulation of blood in the legs and swelling. This is the reason why soldiers who are required to stand still for long periods of time have been known to faint. Elastic bandages around the calf can help prevent this accumulation and can also help provide increased pressure to enable the veins to send blood back up to the heart. For similar reasons, doctors recommend tight stockings for long-haul flights.

Blood pressure may also be measured in the major veins, the heart chambers, arteries to the brain, and the lungs. But these pressures are usually only monitored during surgery or for patients in intensive care since the measurements are invasive. To obtain these pressure measurements, qualified health care workers thread thin tubes, called catheters, into appropriate locations to transmit pressures to external measuring devices.

The heart consists of two pumps—the right side forcing blood through the lungs and the left causing blood to flow through the rest of the body ([link]). Right-heart failure, for example, results in a rise in the pressure in the vena cavae and a drop in pressure in the arteries to the lungs. Left-heart failure results in a rise in the pressure entering the left side of the heart and a drop in aortal pressure. Implications of these and other pressures on flow in the circulatory system will be discussed in more detail in Fluid Dynamics and Its Biological and Medical Applications.

Note:
Two Pumps of the Heart
The heart consists of two pumps—the right side forcing blood through the lungs and the left causing blood to flow through the rest of the body.



Schematic of the circulatory system showing typical pressures. The two pumps in the heart increase pressure and that pressure is reduced as the blood flows through the body. Long-term deviations from these pressures have medical implications discussed in some detail in the Fluid Dynamics and Its Biological and Medical Applications. Only aortal or arterial blood pressure can be measured noninvasively.

Pressure in the Eye

The shape of the eye is maintained by fluid pressure, called **intraocular pressure**, which is normally in the range of 12.0 to 24.0 mm Hg. When the circulation of fluid in the eye is blocked, it can lead to a buildup in pressure, a condition called **glaucoma**. The net pressure can become as great as 85.0 mm Hg, an abnormally large pressure that can permanently damage the optic nerve. To get an idea of the force involved, suppose the back of the eye has an area of $6.0\ \text{cm}^2$, and the net pressure is 85.0 mm Hg. Force is given by $F = \text{PA}$. To get F in newtons, we convert the area to m^2 ($1\ \text{m}^2 = 10^4\ \text{cm}^2$). Then we calculate as follows:
Equation:

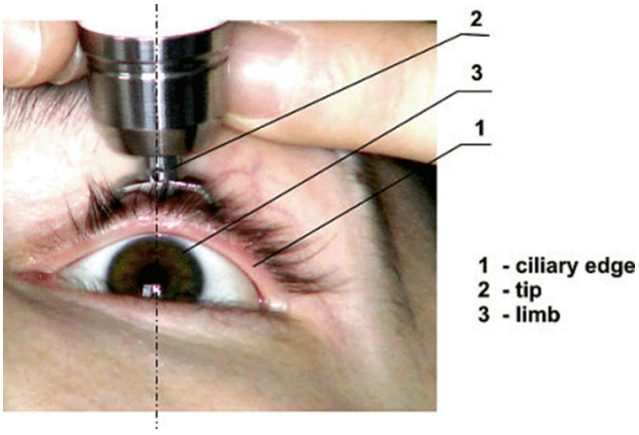
$$F = h\rho gA = \left(85.0 \times 10^{-3}\ \text{m}\right)\left(13.6 \times 10^{3}\ \text{kg/m}^3\right)\left(9.80\ \text{m/s}^2\right)\left(6.0 \times 10^{-4}\ \text{m}^2\right) = 6.8\ \text{N}.$$

Note:
Eye Pressure
The shape of the eye is maintained by fluid pressure, called intraocular pressure. When the circulation of fluid in the eye is blocked, it can lead to a buildup in pressure, a condition called glaucoma. The force is calculated as
Equation:

$$F = h\rho gA = \left(85.0 \times 10^{-3}\ \text{m}\right)\left(13.6 \times 10^{3}\ \text{kg/m}^3\right)\left(9.80\ \text{m/s}^2\right)\left(6.0 \times 10^{-4}\ \text{m}^2\right) = 6.8\ \text{N}.$$

This force is the weight of about a 680-g mass. A mass of 680 g resting on the eye (imagine 1.5 lb resting on your eye) would be sufficient to cause it damage. (A normal force here would be the weight of about 120 g, less than one-quarter of our initial value.)

People over 40 years of age are at greatest risk of developing glaucoma and should have their intraocular pressure tested routinely. Most measurements involve exerting a force on the (anesthetized) eye over some area (a pressure) and observing the eye's response. A noncontact approach uses a puff of air and a measurement is made of the force needed to indent the eye ([link]). If the intraocular pressure is high, the eye will deform less and rebound more vigorously than normal. Excessive intraocular pressures can be detected reliably and sometimes controlled effectively.



The intraocular eye pressure can be read with a tonometer. (credit: DevelopAll at the Wikipedia Project.)

Example:
Calculating Gauge Pressure and Depth: Damage to the Eardrum
Suppose a 3.00-N force can rupture an eardrum. (a) If the eardrum has an area of $1.00\ \text{cm}^2$, calculate the maximum tolerable gauge pressure on the eardrum in newtons per meter squared and convert it to millimeters of mercury. (b) At what depth in freshwater would this person's eardrum rupture, assuming the gauge pressure in the middle ear is zero?
Strategy for (a)
The pressure can be found directly from its definition since we know the force and area. We are looking for the gauge pressure.
Solution for (a)
Equation:

$$P_{\text{g}} = F/A = 3.00\ \text{N}/(1.00 \times 10^{-4}\ \text{m}^2) = 3.00 \times 10^4\ \text{N/m}^2.$$

We now need to convert this to units of mm Hg:
Equation:

$$P_{\text{g}} = 3.0 \times 10^4\ \text{N/m}^2 \left(\frac{1.0\ \text{mm Hg}}{133\ \text{N/m}^2} \right) = 226\ \text{mm Hg}.$$

Strategy for (b)
Here we will use the fact that the water pressure varies linearly with depth h below the surface.
Solution for (b)
$P = h\rho g$ and therefore $h = P/\rho g$. Using the value above for P, we have
Equation:

$$h = \frac{3.0 \times 10^4\ \text{N/m}^2}{(1.00 \times 10^3\ \text{kg/m}^3)(9.80\ \text{m/s}^2)} = 3.06\ \text{m}.$$

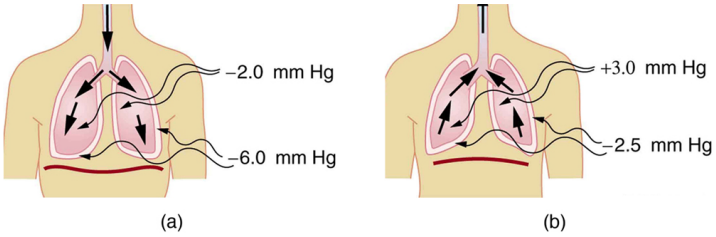
Discussion
Similarly, increased pressure exerted upon the eardrum from the middle ear can arise when an infection causes a fluid buildup.

Pressure Associated with the Lungs

The pressure inside the lungs increases and decreases with each breath. The pressure drops to below atmospheric pressure (negative gauge pressure) when you inhale, causing air to flow into the lungs. It increases above atmospheric pressure (positive gauge pressure) when you exhale, forcing air out.

Lung pressure is controlled by several mechanisms. Muscle action in the diaphragm and rib cage is necessary for inhalation; this muscle action increases the volume of the lungs thereby reducing the pressure within them [link]. Surface tension in the alveoli creates a positive pressure opposing inhalation. (See Cohesion and Adhesion in Liquids: Surface Tension and Capillary Action.) You can exhale without muscle action by letting surface tension in the alveoli create its own positive pressure. Muscle action can add to this positive pressure to produce forced exhalation, such as when you blow up a balloon, blow out a candle, or cough.

The lungs, in fact, would collapse due to the surface tension in the alveoli, if they were not attached to the inside of the chest wall by liquid adhesion. The gauge pressure in the liquid attaching the lungs to the inside of the chest wall is thus negative, ranging from -4 to -8 mm Hg during exhalation and inhalation, respectively. If air is allowed to enter the chest cavity, it breaks the attachment, and one or both lungs may collapse. Suction is applied to the chest cavity of surgery patients and trauma victims to reestablish negative pressure and inflate the lungs.



(a) During inhalation, muscles expand the chest, and the diaphragm moves downward, reducing pressure inside the lungs to less than atmospheric (negative gauge pressure). Pressure between the lungs and chest wall is even lower to overcome the positive pressure created by surface tension in the lungs. (b) During gentle exhalation, the muscles simply relax and surface tension in the alveoli creates a positive pressure inside the lungs, forcing air out. Pressure between the chest wall and lungs remains negative to keep them attached to the chest wall, but it is less negative than during inhalation.

Other Pressures in the Body

Spinal Column and Skull

Normally, there is a 5- to12-mm Hg pressure in the fluid surrounding the brain and filling the spinal column. This cerebrospinal fluid serves many purposes, one of which is to supply flotation to the brain. The buoyant force supplied by the fluid nearly equals the weight of the brain, since their densities are nearly equal. If there is a loss of fluid, the brain rests on the inside of the skull, causing severe headaches, constricted blood flow, and serious damage. Spinal fluid pressure is measured by means of a needle inserted between vertebrae that transmits the pressure to a suitable measuring device.

Bladder Pressure

This bodily pressure is one of which we are often aware. In fact, there is a relationship between our awareness of this pressure and a subsequent increase in it. Bladder pressure climbs steadily from zero to about 25 mm Hg as the bladder fills to its normal capacity of $500\ \text{cm}^3$. This pressure triggers the **micturition reflex**, which stimulates the feeling of needing to urinate. What is more, it also causes muscles around the bladder to contract, raising the pressure to over 100 mm Hg, accentuating the sensation. Coughing, straining, tensing in cold weather, wearing tight clothes, and experiencing simple nervous tension all can increase bladder pressure and trigger this reflex. So can the weight of a pregnant woman's fetus, especially if it is kicking vigorously or pushing down with its head! Bladder pressure can be measured by a catheter or by inserting a needle through the bladder wall and transmitting the pressure to an appropriate measuring device. One hazard of high bladder pressure (sometimes created by an obstruction), is that such pressure can force urine back into the kidneys, causing potentially severe damage.

Pressures in the Skeletal System

These pressures are the largest in the body, due both to the high values of initial force, and the small areas to which this force is applied, such as in the joints.. For example, when a person lifts an object improperly, a force of 5000 N may be created between vertebrae in the spine, and this may be applied to an area as small as $10\ \text{cm}^2$. The pressure created is $P = F/A = (5000\ \text{N})/(10^{-3}\ \text{m}^2) = 5.0 \times 10^6\ \text{N/m}^2$ or about 50 atm! This pressure can damage both the spinal discs (the cartilage between vertebrae), as well as the bony vertebrae themselves. Even under normal circumstances, forces between vertebrae in the spine are large enough to create pressures of several atmospheres. Most causes of excessive pressure in the skeletal system can be avoided by lifting properly and avoiding extreme physical activity. (See Forces and Torques in Muscles and Joints.)

There are many other interesting and medically significant pressures in the body. For example, pressure caused by various muscle actions drives food and waste through the digestive system. Stomach pressure behaves much like bladder pressure and is tied to the sensation of hunger. Pressure in the relaxed esophagus is normally negative because pressure in the chest cavity is normally negative. Positive pressure in the stomach may thus force acid into the esophagus, causing "heartburn." Pressure in the middle ear can result in significant force on the eardrum if it differs greatly from atmospheric pressure, such as while scuba diving. The decrease in external pressure is also noticeable during plane flights (due to a decrease in the weight of air above

relative to that at the Earth’s surface). The Eustachian tubes connect the middle ear to the throat and allow us to equalize pressure in the middle ear to avoid an imbalance of force on the eardrum.

Many pressures in the human body are associated with the flow of fluids. Fluid flow will be discussed in detail in the Fluid Dynamics and Its Biological and Medical Applications.

Section Summary

- Measuring blood pressure is among the most common of all medical examinations.
- The pressures in various parts of the body can be measured and often provide valuable medical indicators.
- The shape of the eye is maintained by fluid pressure, called intraocular pressure.
- When the circulation of fluid in the eye is blocked, it can lead to a buildup in pressure, a condition called glaucoma.
- Some of the other pressures in the body are spinal and skull pressures, bladder pressure, pressures in the skeletal system.

Problems & Exercises

Exercise:

Problem:

During forced exhalation, such as when blowing up a balloon, the diaphragm and chest muscles create a pressure of 60.0 mm Hg between the lungs and chest wall. What force in newtons does this pressure create on the $600\ \text{cm}^2$ surface area of the diaphragm?

Solution:

479 N

Exercise:

Problem:

You can chew through very tough objects with your incisors because they exert a large force on the small area of a pointed tooth. What pressure in pascals can you create by exerting a force of $500\ \text{N}$ with your tooth on an area of $1.00\ \text{mm}^2$?

Exercise:

Problem:

One way to force air into an unconscious person’s lungs is to squeeze on a balloon appropriately connected to the subject. What force must you exert on the balloon with your hands to create a gauge pressure of 4.00 cm water, assuming you squeeze on an effective area of $50.0\ \text{cm}^2$?

Solution:

1.96 N

Exercise:

Problem:

Heroes in movies hide beneath water and breathe through a hollow reed (villains never catch on to this trick). In practice, you cannot inhale in this manner if your lungs are more than 60.0 cm below the surface. What is the maximum negative gauge pressure you can create in your lungs on dry land, assuming you can achieve $-3.00\ \text{cm}$ water pressure with your lungs 60.0 cm below the surface?

Solution:

$-63.0\ \text{cm}\ \text{H}_2\text{O}$

Exercise:

Problem:

Gauge pressure in the fluid surrounding an infant's brain may rise as high as 85.0 mm Hg (5 to 12 mm Hg is normal), creating an outward force large enough to make the skull grow abnormally large. (a) Calculate this outward force in newtons on each side of an infant's skull if the effective area of each side is $70.0\ \text{cm}^2$. (b) What is the net force acting on the skull?

Exercise:

Problem:

A full-term fetus typically has a mass of 3.50 kg. (a) What pressure does the weight of such a fetus create if it rests on the mother's bladder, supported on an area of $90.0\ \text{cm}^2$? (b) Convert this pressure to millimeters of mercury and determine if it alone is great enough to trigger the micturition reflex (it will add to any pressure already existing in the bladder).

Solution:

(a) $3.81 \times 10^3\ \text{N/m}^2$

(b) $28.7\ \text{mm Hg}$, which is sufficient to trigger micturition reflex

Exercise:

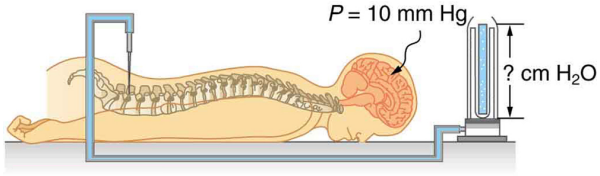
Problem:

If the pressure in the esophagus is $-2.00\ \text{mm Hg}$ while that in the stomach is $+20.0\ \text{mm Hg}$, to what height could stomach fluid rise in the esophagus, assuming a density of 1.10 g/mL? (This movement will not occur if the muscle closing the lower end of the esophagus is working properly.)

Exercise:

Problem:

Pressure in the spinal fluid is measured as shown in [link]. If the pressure in the spinal fluid is 10.0 mm Hg: (a) What is the reading of the water manometer in cm water? (b) What is the reading if the person sits up, placing the top of the fluid 60 cm above the tap? The fluid density is 1.05 g/mL.



A water manometer used to measure pressure in the spinal fluid. The height of the fluid in the manometer is measured relative to the spinal column, and the manometer is open to the atmosphere. The measured pressure will be considerably greater if the person sits up.

Solution:

(a) 13.6 m water

(b) 76.5 cm water

Exercise:

Problem:

Calculate the maximum force in newtons exerted by the blood on an aneurysm, or ballooning, in a major artery, given the maximum blood pressure for this person is 150 mm Hg and the effective area of the aneurysm is $20.0\ \text{cm}^2$. Note that this force is great enough to cause further enlargement and subsequently greater force on the ever-thinner vessel wall.

Exercise:

Problem:

During heavy lifting, a disk between spinal vertebrae is subjected to a 5000-N compressional force. (a) What pressure is created, assuming that the disk has a uniform circular cross section 2.00 cm in radius? (b) What deformation is produced if the disk is 0.800 cm thick and has a Young's modulus of $1.5 \times 10^9\ \text{N/m}^2$?

Solution:

(a) $3.98 \times 10^6\ \text{Pa}$

(b) 2.1×10^{-3} cm

Exercise:

Problem:

When a person sits erect, increasing the vertical position of their brain by 36.0 cm, the heart must continue to pump blood to the brain at the same rate. (a) What is the gain in gravitational potential energy for 100 mL of blood raised 36.0 cm? (b) What is the drop in pressure, neglecting any losses due to friction? (c) Discuss how the gain in gravitational potential energy and the decrease in pressure are related.

Exercise:

Problem:

(a) How high will water rise in a glass capillary tube with a 0.500-mm radius? (b) How much gravitational potential energy does the water gain? (c) Discuss possible sources of this energy.

Solution:

(a) 2.97 cm

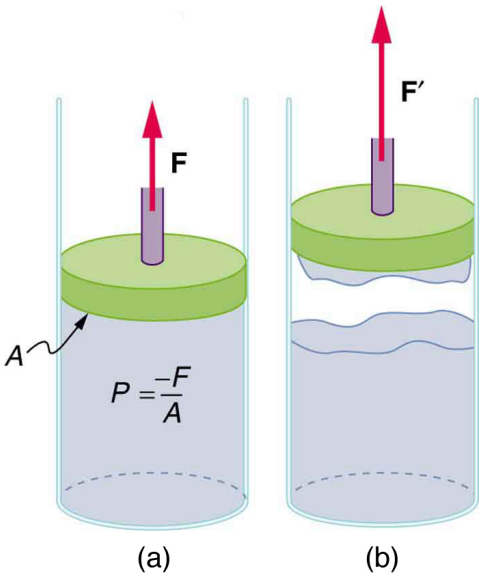
(b) 3.39×10^{-6} J

(c) Work is done by the surface tension force through an effective distance $h/2$ to raise the column of water.

Exercise:

Problem:

A negative pressure of 25.0 atm can sometimes be achieved with the device in [link] before the water separates. (a) To what height could such a negative gauge pressure raise water? (b) How much would a steel wire of the same diameter and length as this capillary stretch if suspended from above?



(a) When the piston is raised, it stretches the liquid slightly, putting it under tension and creating a negative absolute pressure $P = -F/A$ (b) The liquid eventually separates, giving an experimental limit to negative pressure in this liquid.

Exercise:

Problem:

Suppose you hit a steel nail with a 0.500-kg hammer, initially moving at $15.0\ \mathrm{m/s}$ and brought to rest in 2.80 mm. (a) What average force is exerted on the nail? (b) How much is the nail compressed if it is 2.50 mm in diameter and 6.00-cm long? (c) What pressure is created on the 1.00-mm-diameter tip of the nail?

Solution:

(a) $2.01 \times 10^{4}\ \mathrm{N}$

(b) $1.17 \times 10^{-3}\ \mathrm{m}$

(c) $2.56 \times 10^{10}\ \mathrm{N/m^2}$

Exercise:

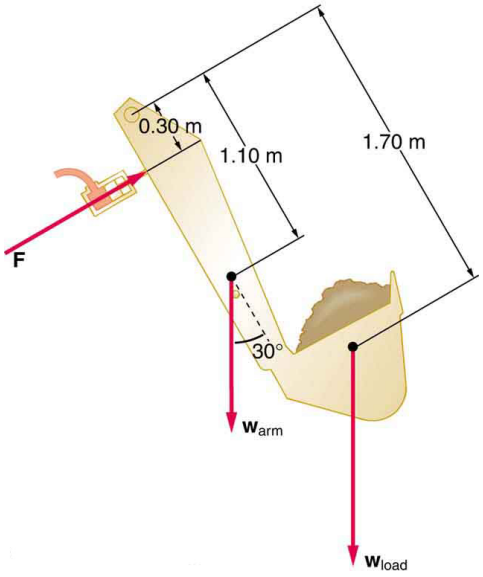
Problem:

Calculate the pressure due to the ocean at the bottom of the Marianas Trench near the Philippines, given its depth is 11.0 km and assuming the density of sea water is constant all the way down. (b) Calculate the percent decrease in volume of sea water due to such a pressure, assuming its bulk modulus is the same as water and is constant. (c) What would be the percent increase in its density? Is the assumption of constant density valid? Will the actual pressure be greater or smaller than that calculated under this assumption?

Exercise:

Problem:

The hydraulic system of a backhoe is used to lift a load as shown in [link]. (a) Calculate the force F the slave cylinder must exert to support the 400-kg load and the 150-kg brace and shovel. (b) What is the pressure in the hydraulic fluid if the slave cylinder is 2.50 cm in diameter? (c) What force would you have to exert on a lever with a mechanical advantage of 5.00 acting on a master cylinder 0.800 cm in diameter to create this pressure?



Hydraulic and mechanical lever systems are used in heavy machinery such as this back hoe.

Solution:

(a) 1.38×10^4 N

(b) 2.81×10^7 N/m^2

(c) 283 N

Exercise:

Problem:

Some miners wish to remove water from a mine shaft. A pipe is lowered to the water 90 m below, and a negative pressure is applied to raise the water. (a) Calculate the pressure needed to raise the water. (b) What is unreasonable about this pressure? (c) What is unreasonable about the premise?

Exercise:

Problem:

You are pumping up a bicycle tire with a hand pump, the piston of which has a 2.00-cm radius.

(a) What force in newtons must you exert to create a pressure of $6.90 \times 10^5\ \mathrm{Pa}$ (b) What is unreasonable about this (a) result? (c) Which premises are unreasonable or inconsistent?

Solution:

(a) 867 N

(b) This is too much force to exert with a hand pump.

(c) The assumed radius of the pump is too large; it would be nearly two inches in diameter —too large for a pump or even a master cylinder. The pressure is reasonable for bicycle tires.

Exercise:

Problem:

Consider a group of people trying to stay afloat after their boat strikes a log in a lake. Construct a problem in which you calculate the number of people that can cling to the log and keep their heads out of the water. Among the variables to be considered are the size and density of the log, and what is needed to keep a person’s head and arms above water without swimming or treading water.

Exercise:

Problem:

The alveoli in emphysema victims are damaged and effectively form larger sacs. Construct a problem in which you calculate the loss of pressure due to surface tension in the alveoli because of their larger average diameters. (Part of the lung’s ability to expel air results from pressure created by surface tension in the alveoli.) Among the things to consider are the normal surface tension of the fluid lining the alveoli, the average alveolar radius in normal individuals and its average in emphysema sufferers.

Glossary

diastolic pressure
: minimum arterial blood pressure; indicator for the fluid balance

glaucoma
: condition caused by the buildup of fluid pressure in the eye

intraocular pressure
: fluid pressure in the eye

micturition reflex
: stimulates the feeling of needing to urinate, triggered by bladder pressure

systolic pressure
: maximum arterial blood pressure; indicator for the blood flow

Structure and Function of Blood Vessels

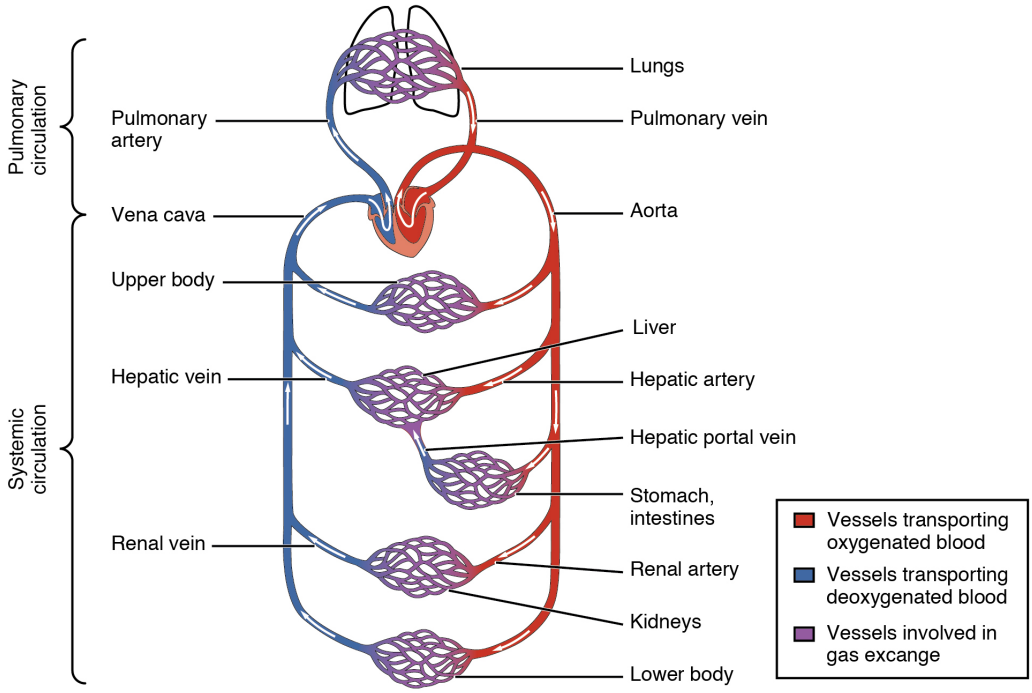
By the end of this section, you will be able to:

- Compare and contrast the three tunics that make up the walls of most blood vessels
- Distinguish between elastic arteries, muscular arteries, and arterioles on the basis of structure, location, and function
- Describe the basic structure of a capillary bed, from the supplying metarteriole to the venule into which it drains
- Explain the structure and function of venous valves in the large veins of the extremities

Blood is carried through the body via blood vessels. An artery is a blood vessel that carries blood away from the heart, where it branches into ever-smaller vessels. Eventually, the smallest arteries, vessels called arterioles, further branch into tiny capillaries, where nutrients and wastes are exchanged, and then combine with other vessels that exit capillaries to form venules, small blood vessels that carry blood to a vein, a larger blood vessel that returns blood to the heart.

Arteries and veins transport blood in two distinct circuits: the systemic circuit and the pulmonary circuit ([link]). Systemic arteries provide blood rich in oxygen to the body's tissues. The blood returned to the heart through systemic veins has less oxygen, since much of the oxygen carried by the arteries has been delivered to the cells. In contrast, in the pulmonary circuit, arteries carry blood low in oxygen exclusively to the lungs for gas exchange. Pulmonary veins then return freshly oxygenated blood from the lungs to the heart to be pumped back out into systemic circulation. Although arteries and veins differ structurally and functionally, they share certain features.

Cardiovascular Circulation



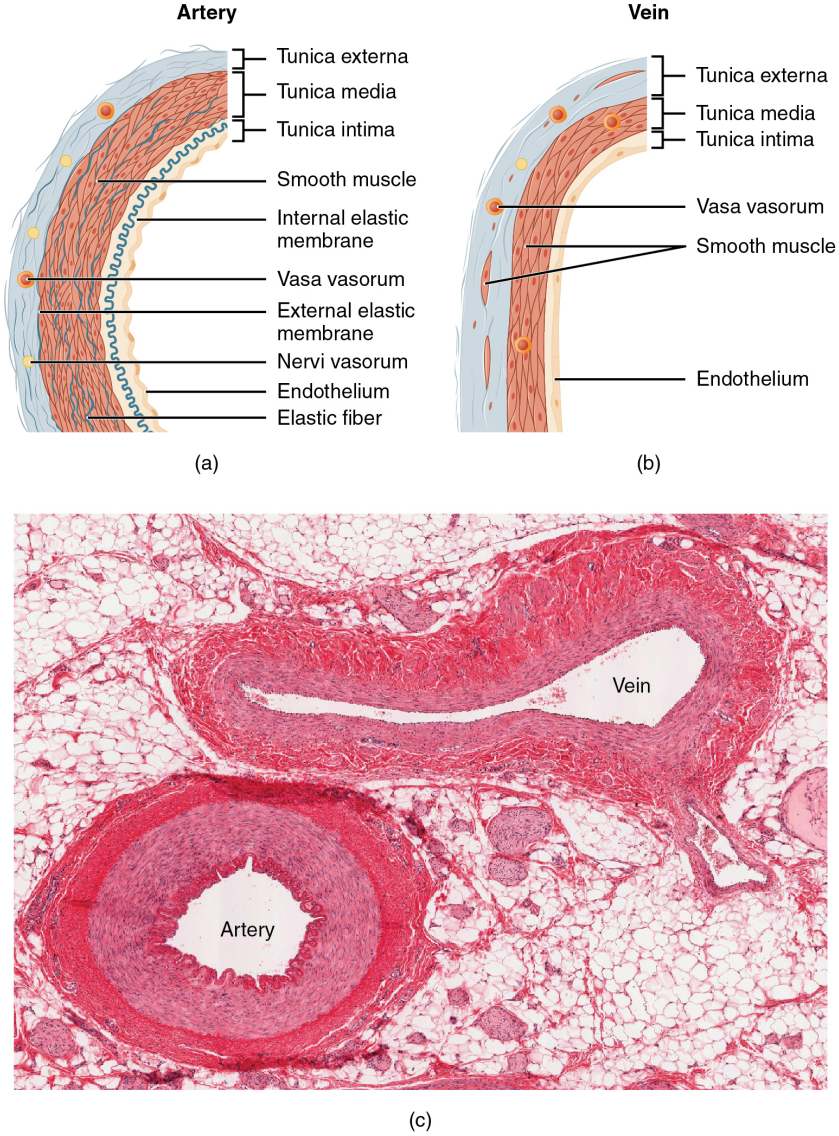
The pulmonary circuit moves blood from the right side of the heart to the lungs and back to the heart. The systemic circuit moves blood from the left side of the heart to the head and body and returns it to the right side of the heart to repeat the cycle. The arrows indicate the direction of blood flow, and the colors show the relative levels of oxygen concentration.

Shared Structures

Different types of blood vessels vary slightly in their structures, but they share the same general features. Arteries and arterioles have thicker walls than veins and venules because they are closer to the heart and receive blood that is surging at a far greater pressure ([[link]]). Each type of vessel has a **lumen**—a hollow passageway through which blood flows. Arteries have smaller lumens than veins, a characteristic that helps to maintain the pressure of blood moving through the system. Together, their thicker walls

and smaller diameters give arterial lumens a more rounded appearance in cross section than the lumens of veins.

Structure of Blood Vessels



(a) Arteries and (b) veins share the same general features, but the walls of arteries are much thicker because of the higher pressure of the blood that flows through them. (c) A micrograph shows the relative differences in thickness. LM × 160. (Micrograph provided by the Regents of the University of Michigan Medical School © 2012)

By the time blood has passed through capillaries and entered venules, the pressure initially exerted upon it by heart contractions has diminished. In other words, in comparison to arteries, venules and veins withstand a much lower pressure from the blood that flows through them. Their walls are considerably thinner and their lumens are correspondingly larger in diameter, allowing more blood to flow with less vessel resistance. In addition, many veins of the body, particularly those of the limbs, contain valves that assist the unidirectional flow of blood toward the heart. This is critical because blood flow becomes sluggish in the extremities, as a result of the lower pressure and the effects of gravity.

The walls of arteries and veins are largely composed of living cells and their products (including collagenous and elastic fibers); the cells require nourishment and produce waste. Since blood passes through the larger vessels relatively quickly, there is limited opportunity for blood in the lumen of the vessel to provide nourishment to or remove waste from the vessel's cells. Further, the walls of the larger vessels are too thick for nutrients to diffuse through to all of the cells. Larger arteries and veins contain small blood vessels within their walls known as the **vasa vasorum**—literally "vessels of the vessel"—to provide them with this critical exchange. Since the pressure within arteries is relatively high, the vasa vasorum must function in the outer layers of the vessel (see [link]) or the pressure exerted by the blood passing through the vessel would collapse it, preventing any exchange from occurring. The lower pressure within veins allows the vasa vasorum to be located closer to the lumen. The restriction of the vasa vasorum to the outer layers of arteries is thought to be one reason that arterial diseases are more common than venous diseases, since its location makes it more difficult to nourish the cells of the arteries and remove waste products. There are also minute nerves within the walls of both types of vessels that control the contraction and dilation of smooth muscle. These minute nerves are known as the nervi vasorum.

Both arteries and veins have the same three distinct tissue layers, called tunics (from the Latin term tunica), for the garments first worn by ancient Romans; the term tunic is also used for some modern garments. From the

most interior layer to the outer, these tunics are the tunica intima, the tunica media, and the tunica externa (see [link]). [link] compares and contrasts the tunics of the arteries and veins.

Comparison of Tunics in Arteries and Veins		
	Arteries	**Veins**
General appearance	Thick walls with small lumens Generally appear rounded	Thin walls with large lumens Generally appear flattened
Tunica intima	Endothelium usually appears wavy due to constriction of smooth muscle Internal elastic membrane present in larger vessels	Endothelium appears smooth Internal elastic membrane absent

Comparison of Tunics in Arteries and Veins		
	Arteries	**Veins**
Tunica media	Normally the thickest layer in arteries Smooth muscle cells and elastic fibers predominate (the proportions of these vary with distance from the heart) External elastic membrane present in larger vessels	Normally thinner than the tunica externa Smooth muscle cells and collagenous fibers predominate Nervi vasorum and vasa vasorum present External elastic membrane absent
Tunica externa	Normally thinner than the tunica media in all but the largest arteries Collagenous and elastic fibers Nervi vasorum and vasa vasorum present	Normally the thickest layer in veins Collagenous and smooth fibers predominate Some smooth muscle fibers Nervi vasorum and vasa vasorum present

Tunica Intima

The **tunica intima** (also called the tunica interna) is composed of epithelial and connective tissue layers. Lining the tunica intima is the specialized simple squamous epithelium called the endothelium, which is continuous throughout the entire vascular system, including the lining of the chambers of the heart. Damage to this endothelial lining and exposure of blood to the collagenous fibers beneath is one of the primary causes of clot formation. Until recently, the endothelium was viewed simply as the boundary between the blood in the lumen and the walls of the vessels. Recent studies, however, have shown that it is physiologically critical to such activities as helping to regulate capillary exchange and altering blood flow. The endothelium releases local chemicals called endothelins that can constrict the smooth muscle within the walls of the vessel to increase blood pressure. Uncompensated overproduction of endothelins may contribute to hypertension (high blood pressure) and cardiovascular disease.

Next to the endothelium is the basement membrane, or basal lamina, that effectively binds the endothelium to the connective tissue. The basement membrane provides strength while maintaining flexibility, and it is permeable, allowing materials to pass through it. The thin outer layer of the tunica intima contains a small amount of areolar connective tissue that consists primarily of elastic fibers to provide the vessel with additional flexibility; it also contains some collagenous fibers to provide additional strength.

In larger arteries, there is also a thick, distinct layer of elastic fibers known as the **internal elastic membrane** (also called the internal elastic lamina) at the boundary with the tunica media. Like the other components of the tunica intima, the internal elastic membrane provides structure while allowing the vessel to stretch. It is permeated with small openings that allow exchange of materials between the tunics. The internal elastic membrane is not apparent in veins. In addition, many veins, particularly in the lower limbs, contain valves formed by sections of thickened endothelium that are reinforced with connective tissue, extending into the lumen.

Under the microscope, the lumen and the entire tunica intima of a vein will appear smooth, whereas those of an artery will normally appear wavy because of the partial constriction of the smooth muscle in the tunica media, the next layer of blood vessel walls.

Tunica Media

The **tunica media** is the substantial middle layer of the vessel wall (see [link]). It is generally the thickest layer in arteries, and it is much thicker in arteries than it is in veins. The tunica media consists of layers of smooth muscle supported by connective tissue that is primarily made up of elastic fibers, most of which are arranged in circular sheets. Toward the outer portion of the tunic, there are also layers of longitudinal muscle. Contraction and relaxation of the circular muscles decrease and increase the diameter of the vessel lumen, respectively. Specifically in arteries, **vasoconstriction** decreases blood flow as the smooth muscle in the walls of the tunica media contracts, making the lumen narrower and increasing blood pressure. Similarly, **vasodilation** increases blood flow as the smooth muscle relaxes, allowing the lumen to widen and blood pressure to drop. Both vasoconstriction and vasodilation are regulated in part by small vascular nerves, known as **nervi vasorum**, or “nerves of the vessel,” that run within the walls of blood vessels. These are generally all sympathetic fibers, although some trigger vasodilation and others induce vasoconstriction, depending upon the nature of the neurotransmitter and receptors located on the target cell. Parasympathetic stimulation does trigger vasodilation as well as erection during sexual arousal in the external genitalia of both sexes. Nervous control over vessels tends to be more generalized than the specific targeting of individual blood vessels. Local controls, discussed later, account for this phenomenon. (Seek additional content for more information on these dynamic aspects of the autonomic nervous system.) Hormones and local chemicals also control blood vessels. Together, these neural and chemical mechanisms reduce or increase blood flow in response to changing body conditions, from exercise to hydration. Regulation of both blood flow and blood pressure is discussed in detail later in this chapter.

The smooth muscle layers of the tunica media are supported by a framework of collagenous fibers that also binds the tunica media to the inner and outer tunics. Along with the collagenous fibers are large numbers of elastic fibers that appear as wavy lines in prepared slides. Separating the tunica media from the outer tunica externa in larger arteries is the **external elastic membrane** (also called the external elastic lamina), which also appears wavy in slides. This structure is not usually seen in smaller arteries, nor is it seen in veins.

Tunica Externa

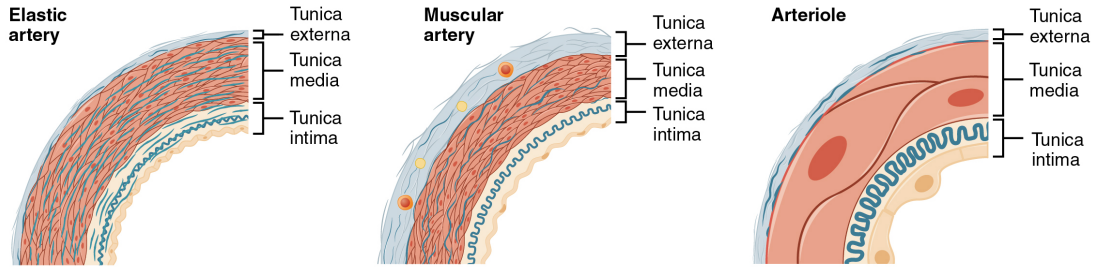
The outer tunic, the **tunica externa** (also called the tunica adventitia), is a substantial sheath of connective tissue composed primarily of collagenous fibers. Some bands of elastic fibers are found here as well. The tunica externa in veins also contains groups of smooth muscle fibers. This is normally the thickest tunic in veins and may be thicker than the tunica media in some larger arteries. The outer layers of the tunica externa are not distinct but rather blend with the surrounding connective tissue outside the vessel, helping to hold the vessel in relative position. If you are able to palpate some of the superficial veins on your upper limbs and try to move them, you will find that the tunica externa prevents this. If the tunica externa did not hold the vessel in place, any movement would likely result in disruption of blood flow.

Arteries

An **artery** is a blood vessel that conducts blood away from the heart. All arteries have relatively thick walls that can withstand the high pressure of blood ejected from the heart. However, those close to the heart have the thickest walls, containing a high percentage of elastic fibers in all three of their tunics. This type of artery is known as an **elastic artery** ([link]). Vessels larger than 10 mm in diameter are typically elastic. Their abundant elastic fibers allow them to expand, as blood pumped from the ventricles passes through them, and then to recoil after the surge has passed. If artery walls were rigid and unable to expand and recoil, their resistance to blood

flow would greatly increase and blood pressure would rise to even higher levels, which would in turn require the heart to pump harder to increase the volume of blood expelled by each pump (the stroke volume) and maintain adequate pressure and flow. Artery walls would have to become even thicker in response to this increased pressure. The elastic recoil of the vascular wall helps to maintain the pressure gradient that drives the blood through the arterial system. An elastic artery is also known as a conducting artery, because the large diameter of the lumen enables it to accept a large volume of blood from the heart and conduct it to smaller branches.

Types of Arteries and Arterioles



Comparison of the walls of an elastic artery, a muscular artery, and an arteriole is shown. In terms of scale, the diameter of an arteriole is measured in micrometers compared to millimeters for elastic and muscular arteries.

Farther from the heart, where the surge of blood has dampened, the percentage of elastic fibers in an artery's tunica intima decreases and the amount of smooth muscle in its tunica media increases. The artery at this point is described as a **muscular artery**. The diameter of muscular arteries typically ranges from 0.1 mm to 10 mm. Their thick tunica media allows muscular arteries to play a leading role in vasoconstriction. In contrast, their decreased quantity of elastic fibers limits their ability to expand. Fortunately, because the blood pressure has eased by the time it reaches these more distant vessels, elasticity has become less important.

Notice that although the distinctions between elastic and muscular arteries are important, there is no "line of demarcation" where an elastic artery suddenly becomes muscular. Rather, there is a gradual transition as the

vascular tree repeatedly branches. In turn, muscular arteries branch to distribute blood to the vast network of arterioles. For this reason, a muscular artery is also known as a distributing artery.

Arterioles

An **arteriole** is a very small artery that leads to a capillary. Arterioles have the same three tunics as the larger vessels, but the thickness of each is greatly diminished. The critical endothelial lining of the tunica intima is intact. The tunica media is restricted to one or two smooth muscle cell layers in thickness. The tunica externa remains but is very thin (see [link]).

With a lumen averaging 30 micrometers or less in diameter, arterioles are critical in slowing down—or resisting—blood flow and, thus, causing a substantial drop in blood pressure. Because of this, you may see them referred to as resistance vessels. The muscle fibers in arterioles are normally slightly contracted, causing arterioles to maintain a consistent muscle tone—in this case referred to as vascular tone—in a similar manner to the muscular tone of skeletal muscle. In reality, all blood vessels exhibit vascular tone due to the partial contraction of smooth muscle. The importance of the arterioles is that they will be the primary site of both resistance and regulation of blood pressure. The precise diameter of the lumen of an arteriole at any given moment is determined by neural and chemical controls, and vasoconstriction and vasodilation in the arterioles are the primary mechanisms for distribution of blood flow.

Capillaries

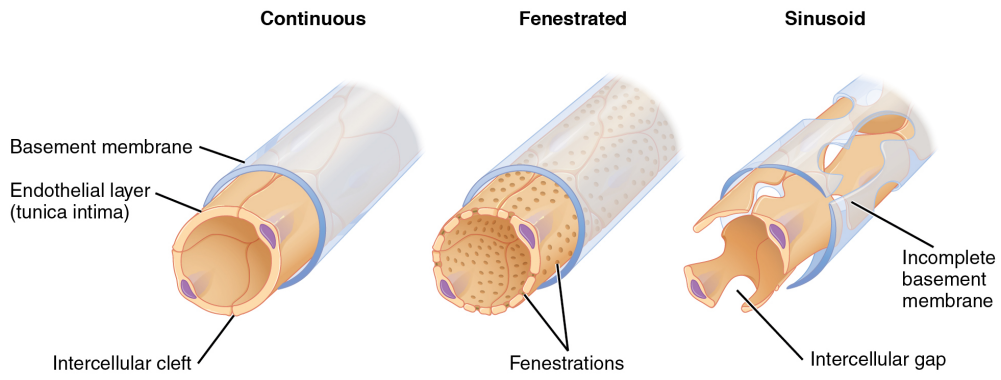
A **capillary** is a microscopic channel that supplies blood to the tissues themselves, a process called **perfusion**. Exchange of gases and other substances occurs in the capillaries between the blood and the surrounding cells and their tissue fluid (interstitial fluid). The diameter of a capillary lumen ranges from 5–10 micrometers; the smallest are just barely wide enough for an erythrocyte to squeeze through. Flow through capillaries is often described as **microcirculation**.

The wall of a capillary consists of the endothelial layer surrounded by a basement membrane with occasional smooth muscle fibers. There is some variation in wall structure: In a large capillary, several endothelial cells bordering each other may line the lumen; in a small capillary, there may be only a single cell layer that wraps around to contact itself.

For capillaries to function, their walls must be leaky, allowing substances to pass through. There are three major types of capillaries, which differ according to their degree of “leakiness:” continuous, fenestrated, and sinusoid capillaries ([link]).

Continuous Capillaries

The most common type of capillary, the **continuous capillary**, is found in almost all vascularized tissues. Continuous capillaries are characterized by a complete endothelial lining with tight junctions between endothelial cells. Although a tight junction is usually impermeable and only allows for the passage of water and ions, they are often incomplete in capillaries, leaving intercellular clefts that allow for exchange of water and other very small molecules between the blood plasma and the interstitial fluid. Substances that can pass between cells include metabolic products, such as glucose, water, and small hydrophobic molecules like gases and hormones, as well as various leukocytes. Continuous capillaries not associated with the brain are rich in transport vesicles, contributing to either endocytosis or exocytosis. Those in the brain are part of the blood-brain barrier. Here, there are tight junctions and no intercellular clefts, plus a thick basement membrane and astrocyte extensions called end feet; these structures combine to prevent the movement of nearly all substances.
Types of Capillaries



The three major types of capillaries: continuous, fenestrated, and sinusoid.

Fenestrated Capillaries

A **fenestrated capillary** is one that has pores (or fenestrations) in addition to tight junctions in the endothelial lining. These make the capillary permeable to larger molecules. The number of fenestrations and their degree of permeability vary, however, according to their location. Fenestrated capillaries are common in the small intestine, which is the primary site of nutrient absorption, as well as in the kidneys, which filter the blood. They are also found in the choroid plexus of the brain and many endocrine structures, including the hypothalamus, pituitary, pineal, and thyroid glands.

Sinusoid Capillaries

A **sinusoid capillary** (or sinusoid) is the least common type of capillary. Sinusoid capillaries are flattened, and they have extensive intercellular gaps and incomplete basement membranes, in addition to intercellular clefts and fenestrations. This gives them an appearance not unlike Swiss cheese. These very large openings allow for the passage of the largest molecules, including plasma proteins and even cells. Blood flow through sinusoids is

very slow, allowing more time for exchange of gases, nutrients, and wastes. Sinusoids are found in the liver and spleen, bone marrow, lymph nodes (where they carry lymph, not blood), and many endocrine glands including the pituitary and adrenal glands. Without these specialized capillaries, these organs would not be able to provide their myriad of functions. For example, when bone marrow forms new blood cells, the cells must enter the blood supply and can only do so through the large openings of a sinusoid capillary; they cannot pass through the small openings of continuous or fenestrated capillaries. The liver also requires extensive specialized sinusoid capillaries in order to process the materials brought to it by the hepatic portal vein from both the digestive tract and spleen, and to release plasma proteins into circulation.

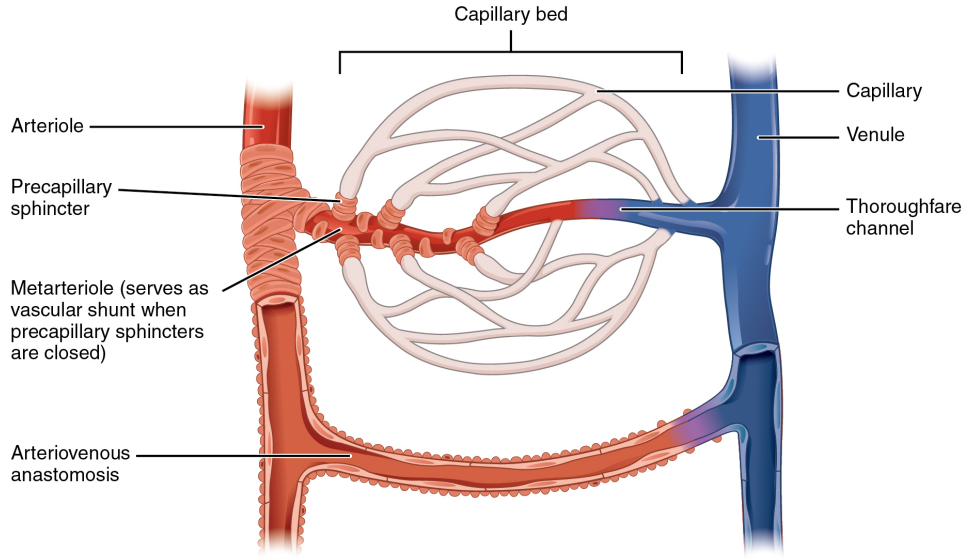
Metarterioles and Capillary Beds

A **metarteriole** is a type of vessel that has structural characteristics of both an arteriole and a capillary. Slightly larger than the typical capillary, the smooth muscle of the tunica media of the metarteriole is not continuous but forms rings of smooth muscle (sphincters) prior to the entrance to the capillaries. Each metarteriole arises from a terminal arteriole and branches to supply blood to a **capillary bed** that may consist of 10–100 capillaries.

The **precapillary sphincters**, circular smooth muscle cells that surround the capillary at its origin with the metarteriole, tightly regulate the flow of blood from a metarteriole to the capillaries it supplies. Their function is critical: If all of the capillary beds in the body were to open simultaneously, they would collectively hold every drop of blood in the body and there would be none in the arteries, arterioles, venules, veins, or the heart itself. Normally, the precapillary sphincters are closed. When the surrounding tissues need oxygen and have excess waste products, the precapillary sphincters open, allowing blood to flow through and exchange to occur before closing once more ([link]). If all of the precapillary sphincters in a capillary bed are closed, blood will flow from the metarteriole directly into a **thoroughfare channel** and then into the venous circulation, bypassing the capillary bed entirely. This creates what is known as a **vascular shunt**. In addition, an **arteriovenous anastomosis** may bypass the capillary bed and lead directly to the venous system.

Although you might expect blood flow through a capillary bed to be smooth, in reality, it moves with an irregular, pulsating flow. This pattern is called **vasomotion** and is regulated by chemical signals that are triggered in response to changes in internal conditions, such as oxygen, carbon dioxide, hydrogen ion, and lactic acid levels. For example, during strenuous exercise when oxygen levels decrease and carbon dioxide, hydrogen ion, and lactic acid levels all increase, the capillary beds in skeletal muscle are open, as they would be in the digestive system when nutrients are present in the digestive tract. During sleep or rest periods, vessels in both areas are largely closed; they open only occasionally to allow oxygen and nutrient supplies to travel to the tissues to maintain basic life processes.

Capillary Bed



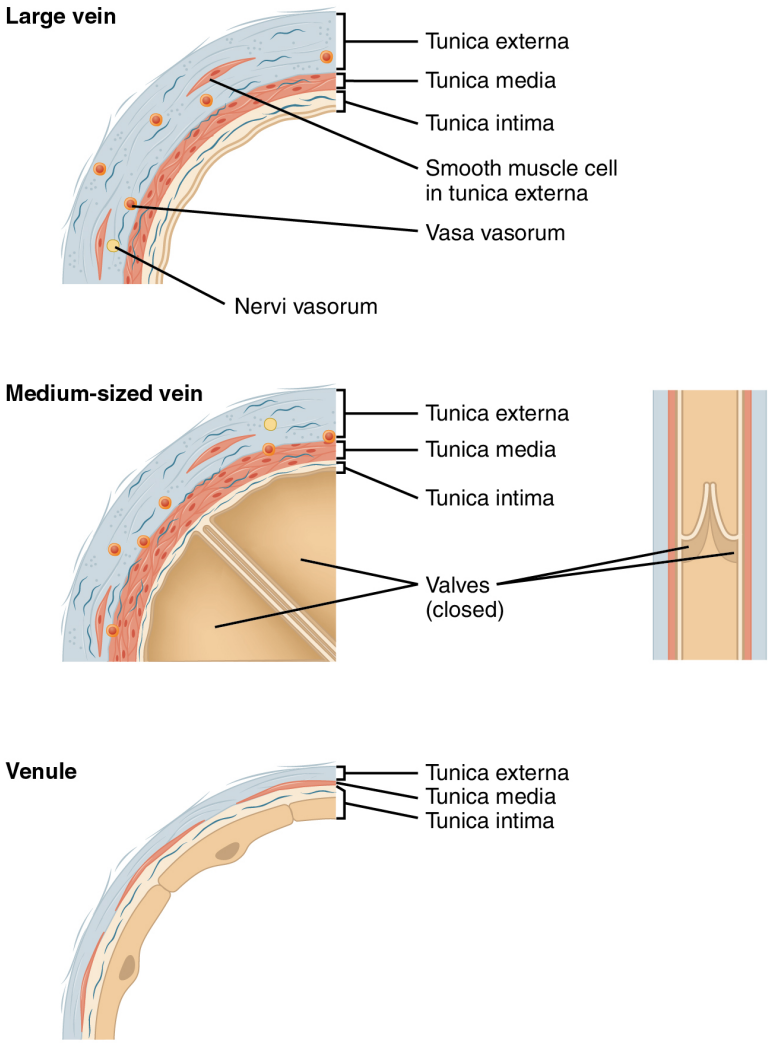
In a capillary bed, arterioles give rise to metarterioles. Precapillary sphincters located at the junction of a metarteriole with a capillary regulate blood flow. A thoroughfare channel connects the metarteriole to a venule. An arteriovenous anastomosis, which directly connects the arteriole with the venule, is shown at the bottom.

Venules

A **venule** is an extremely small vein, generally 8–100 micrometers in diameter. Postcapillary venules join multiple capillaries exiting from a capillary bed. Multiple venules join to form veins. The walls of venules consist of endothelium, a thin middle layer with a few muscle cells and elastic fibers, plus an outer layer of connective tissue fibers that constitute a very thin tunica externa ([link]). Venules as well as capillaries are the primary sites of emigration or diapedesis, in which the white blood cells adhere to the endothelial lining of the vessels and then squeeze through adjacent cells to enter the tissue fluid.

Veins

A **vein** is a blood vessel that conducts blood toward the heart. Compared to arteries, veins are thin-walled vessels with large and irregular lumens (see [link]). Because they are low-pressure vessels, larger veins are commonly equipped with valves that promote the unidirectional flow of blood toward the heart and prevent backflow toward the capillaries caused by the inherent low blood pressure in veins as well as the pull of gravity. [link] compares the features of arteries and veins.
Comparison of Veins and Venules



Many veins have valves to prevent back flow of blood, whereas venules do not. In terms of scale, the diameter of a venule is measured in micrometers compared to millimeters for veins.

Comparison of Arteries and Veins		
	Arteries	**Veins**
Direction of blood flow	Conducts blood away from the heart	Conducts blood toward the heart
General appearance	Rounded	Irregular, often collapsed
Pressure	High	Low
Wall thickness	Thick	Thin
Relative oxygen concentration	Higher in systemic arteries Lower in pulmonary arteries	Lower in systemic veins Higher in pulmonary veins
Valves	Not present	Present most commonly in limbs and in veins inferior to the heart

Note:
Disorders of the…
Cardiovascular System: Edema and Varicose Veins
Despite the presence of valves and the contributions of other anatomical and physiological adaptations we will cover shortly, over the course of a day, some blood will inevitably pool, especially in the lower limbs, due to the pull of gravity. Any blood that accumulates in a vein will increase the

pressure within it, which can then be reflected back into the smaller veins, venules, and eventually even the capillaries. Increased pressure will promote the flow of fluids out of the capillaries and into the interstitial fluid. The presence of excess tissue fluid around the cells leads to a condition called edema.

Most people experience a daily accumulation of tissue fluid, especially if they spend much of their work life on their feet (like most health professionals). However, clinical edema goes beyond normal swelling and requires medical treatment. Edema has many potential causes, including hypertension and heart failure, severe protein deficiency, renal failure, and many others. In order to treat edema, which is a sign rather than a discrete disorder, the underlying cause must be diagnosed and alleviated.

Varicose Veins



Varicose veins are commonly found in the lower limbs. (credit: Thomas Kriese)

Edema may be accompanied by varicose veins, especially in the superficial veins of the legs ([link]). This disorder arises when defective valves allow blood to accumulate within the veins, causing them to distend, twist, and become visible on the surface of the integument. Varicose veins may occur in both sexes, but are more common in women and are often related to pregnancy. More than simple cosmetic blemishes, varicose veins are often painful and sometimes itchy or throbbing. Without treatment, they tend to grow worse over time. The use of support hose, as well as elevating the feet and legs whenever possible, may be helpful in alleviating this condition. Laser surgery and interventional radiologic procedures can reduce the size and severity of varicose veins. Severe cases may require conventional surgery to remove the damaged vessels. As there are typically redundant circulation patterns, that is, anastomoses, for the smaller and more superficial veins, removal does not typically impair the circulation. There is evidence that patients with varicose veins suffer a greater risk of developing a thrombus or clot.

Veins as Blood Reservoirs

In addition to their primary function of returning blood to the heart, veins may be considered blood reservoirs, since systemic veins contain approximately 64 percent of the blood volume at any given time ([link]). Their ability to hold this much blood is due to their high **capacitance**, that is, their capacity to distend (expand) readily to store a high volume of blood, even at a low pressure. The large lumens and relatively thin walls of veins make them far more distensible than arteries; thus, they are said to be **capacitance vessels**.

Distribution of Blood Flow

Systemic circulation 84%	Systemic veins 64%	Large veins 18%
		Large venous networks (liver, bone marrow, and integument) 21%
		Venules and medium-sized veins 25%
	Systemic arteries 13%	Arterioles 2%
		Muscular arteries 5%
		Elastic arteries 4%
		Aorta 2%
	Systemic capillaries 7%	Systemic capillaries 7%
Pulmonary circulation 9%	Pulmonary veins 4%	
	Pulmonary capillaries 2%	
	Pulmonary arteries 3%	
Heart 7%		

When blood flow needs to be redistributed to other portions of the body, the vasomotor center located in the medulla oblongata sends sympathetic stimulation to the smooth muscles in the walls of the veins, causing constriction—or in this case, venoconstriction. Less dramatic than the vasoconstriction seen in smaller arteries and arterioles, venoconstriction may be likened to a “stiffening” of the vessel wall. This increases pressure on the blood within the veins, speeding its return to the heart. As you will note in [link], approximately 21 percent of the venous blood is located in venous networks within the liver, bone marrow, and integument. This volume of blood is referred to as **venous reserve**. Through venoconstriction, this “reserve” volume of blood can get back to the heart more quickly for redistribution to other parts of the circulation.

Note:
Career Connection
Vascular Surgeons and Technicians

Vascular surgery is a specialty in which the physician deals primarily with diseases of the vascular portion of the cardiovascular system. This includes repair and replacement of diseased or damaged vessels, removal of plaque from vessels, minimally invasive procedures including the insertion of venous catheters, and traditional surgery. Following completion of medical school, the physician generally completes a 5-year surgical residency followed by an additional 1 to 2 years of vascular specialty training. In the United States, most vascular surgeons are members of the Society of Vascular Surgery.

Vascular technicians are specialists in imaging technologies that provide information on the health of the vascular system. They may also assist physicians in treating disorders involving the arteries and veins. This profession often overlaps with cardiovascular technology, which would also include treatments involving the heart. Although recognized by the American Medical Association, there are currently no licensing requirements for vascular technicians, and licensing is voluntary. Vascular technicians typically have an Associate's degree or certificate, involving 18 months to 2 years of training. The United States Bureau of Labor projects this profession to grow by 29 percent from 2010 to 2020.

Note:



Visit this site to learn more about vascular surgery.

Note:



Visit this site to learn more about vascular technicians.

Chapter Review

Blood pumped by the heart flows through a series of vessels known as arteries, arterioles, capillaries, venules, and veins before returning to the heart. Arteries transport blood away from the heart and branch into smaller vessels, forming arterioles. Arterioles distribute blood to capillary beds, the sites of exchange with the body tissues. Capillaries lead back to small vessels known as venules that flow into the larger veins and eventually back to the heart.

The arterial system is a relatively high-pressure system, so arteries have thick walls that appear round in cross section. The venous system is a lower-pressure system, containing veins that have larger lumens and thinner walls. They often appear flattened. Arteries, arterioles, venules, and veins are composed of three tunics known as the tunica intima, tunica media, and tunica externa. Capillaries have only a tunica intima layer. The tunica intima is a thin layer composed of a simple squamous epithelium known as endothelium and a small amount of connective tissue. The tunica media is a thicker area composed of variable amounts of smooth muscle and connective tissue. It is the thickest layer in all but the largest arteries. The tunica externa is primarily a layer of connective tissue, although in veins, it also contains some smooth muscle. Blood flow through vessels can be dramatically influenced by vasoconstriction and vasodilation in their walls.

Review Questions

Exercise:

Problem: The endothelium is found in the _____.

a. tunica intima
b. tunica media
c. tunica externa
d. lumen

Solution:

A

Exercise:

Problem: Nervi vasorum control _____.

a. vasoconstriction
b. vasodilation
c. capillary permeability
d. both vasoconstriction and vasodilation

Solution:

D

Exercise:

Problem:

Closer to the heart, arteries would be expected to have a higher percentage of _____.

a. endothelium
b. smooth muscle fibers
c. elastic fibers
d. collagenous fibers

Solution:

C

Exercise:

Problem: Which of the following best describes veins?

a. thick walled, small lumens, low pressure, lack valves
b. thin walled, large lumens, low pressure, have valves
c. thin walled, small lumens, high pressure, have valves
d. thick walled, large lumens, high pressure, lack valves

Solution:

B

Exercise:

Problem:

An especially leaky type of capillary found in the liver and certain other tissues is called a _____.

a. capillary bed
b. fenestrated capillary
c. sinusoid capillary
d. metarteriole

Solution:

C

Critical Thinking Questions

Exercise:

Problem: Arterioles are often referred to as resistance vessels. Why?

Solution:

Arterioles receive blood from arteries, which are vessels with a much larger lumen. As their own lumen averages just 30 micrometers or less, arterioles are critical in slowing down—or resisting—blood flow. The arterioles can also constrict or dilate, which varies their resistance, to help distribute blood flow to the tissues.

Exercise:

Problem:

Cocaine use causes vasoconstriction. Is this likely to increase or decrease blood pressure, and why?

Solution:

Vasoconstriction causes the lumens of blood vessels to narrow. This increases the pressure of the blood flowing within the vessel.

Exercise:

Problem:

A blood vessel with a few smooth muscle fibers and connective tissue, and only a very thin tunica externa conducts blood toward the heart. What type of vessel is this?

Solution:

This is a venule.

Glossary

arteriole
: (also, resistance vessel) very small artery that leads to a capillary

arteriovenous anastomosis
: short vessel connecting an arteriole directly to a venule and bypassing the capillary beds

artery
: blood vessel that conducts blood away from the heart; may be a conducting or distributing vessel

capacitance
: ability of a vein to distend and store blood

capacitance vessels
: veins

capillary
: smallest of blood vessels where physical exchange occurs between the blood and tissue cells surrounded by interstitial fluid

capillary bed
: network of 10–100 capillaries connecting arterioles to venules

continuous capillary
: most common type of capillary, found in virtually all tissues except epithelia and cartilage; contains very small gaps in the endothelial lining that permit exchange

elastic artery
: (also, conducting artery) artery with abundant elastic fibers located closer to the heart, which maintains the pressure gradient and conducts blood to smaller branches

external elastic membrane
: membrane composed of elastic fibers that separates the tunica media from the tunica externa; seen in larger arteries

fenestrated capillary
: type of capillary with pores or fenestrations in the endothelium that allow for rapid passage of certain small materials

internal elastic membrane
: membrane composed of elastic fibers that separates the tunica intima from the tunica media; seen in larger arteries

lumen
: interior of a tubular structure such as a blood vessel or a portion of the alimentary canal through which blood, chyme, or other substances travel

metarteriole
: short vessel arising from a terminal arteriole that branches to supply a capillary bed

microcirculation
: blood flow through the capillaries

muscular artery
: (also, distributing artery) artery with abundant smooth muscle in the tunica media that branches to distribute blood to the arteriole network

nervi vasorum
: small nerve fibers found in arteries and veins that trigger contraction of the smooth muscle in their walls

perfusion
: distribution of blood into the capillaries so the tissues can be supplied

precapillary sphincters
: circular rings of smooth muscle that surround the entrance to a capillary and regulate blood flow into that capillary

sinusoid capillary
: rarest type of capillary, which has extremely large intercellular gaps in the basement membrane in addition to clefts and fenestrations; found in areas such as the bone marrow and liver where passage of large molecules occurs

thoroughfare channel

continuation of the metarteriole that enables blood to bypass a capillary bed and flow directly into a venule, creating a vascular shunt

tunica externa
: (also, tunica adventitia) outermost layer or tunic of a vessel (except capillaries)

tunica intima
: (also, tunica interna) innermost lining or tunic of a vessel

tunica media
: middle layer or tunic of a vessel (except capillaries)

vasa vasorum
: small blood vessels located within the walls or tunics of larger vessels that supply nourishment to and remove wastes from the cells of the vessels

vascular shunt
: continuation of the metarteriole and thoroughfare channel that allows blood to bypass the capillary beds to flow directly from the arterial to the venous circulation

vasoconstriction
: constriction of the smooth muscle of a blood vessel, resulting in a decreased vascular diameter

vasodilation
: relaxation of the smooth muscle in the wall of a blood vessel, resulting in an increased vascular diameter

vasomotion
: irregular, pulsating flow of blood through capillaries and related structures

vein
: blood vessel that conducts blood toward the heart

venous reserve
: volume of blood contained within systemic veins in the integument, bone marrow, and liver that can be returned to the heart for circulation, if needed

venule
: small vessel leading from the capillaries to veins

An Overview of Blood
By the end of this section, you will be able to:

- Identify the primary functions of blood in transportation, defense, and maintenance of homeostasis
- Name the fluid component of blood and the three major types of formed elements, and identify their relative proportions in a blood sample
- Discuss the unique physical characteristics of blood
- Identify the composition of blood plasma, including its most important solutes and plasma proteins

Recall that **blood** is a connective tissue. Like all connective tissues, it is made up of cellular elements and an extracellular matrix. The cellular elements—referred to as the **formed elements**—include **red blood cells (RBCs)**, **white blood cells (WBCs)**, and cell fragments called **platelets**. The extracellular matrix, called **plasma**, makes blood unique among connective tissues because it is fluid. This fluid, which is mostly water, perpetually suspends the formed elements and enables them to circulate throughout the body within the cardiovascular system.

Functions of Blood

The primary function of blood is to deliver oxygen and nutrients to and remove wastes from body cells, but that is only the beginning of the story. The specific functions of blood also include defense, distribution of heat, and maintenance of homeostasis.

Transportation

Nutrients from the foods you eat are absorbed in the digestive tract. Most of these travel in the bloodstream directly to the liver, where they are processed and released back into the bloodstream for delivery to body cells. Oxygen from the air you breathe diffuses into the blood, which moves from the lungs to the heart, which then pumps it out to the rest of the body. Moreover, endocrine glands scattered throughout the body release their

products, called hormones, into the bloodstream, which carries them to distant target cells. Blood also picks up cellular wastes and byproducts, and transports them to various organs for removal. For instance, blood moves carbon dioxide to the lungs for exhalation from the body, and various waste products are transported to the kidneys and liver for excretion from the body in the form of urine or bile.

Defense

Many types of WBCs protect the body from external threats, such as disease-causing bacteria that have entered the bloodstream in a wound. Other WBCs seek out and destroy internal threats, such as cells with mutated DNA that could multiply to become cancerous, or body cells infected with viruses.

When damage to the vessels results in bleeding, blood platelets and certain proteins dissolved in the plasma, the fluid portion of the blood, interact to block the ruptured areas of the blood vessels involved. This protects the body from further blood loss.

Maintenance of Homeostasis

Recall that body temperature is regulated via a classic negative-feedback loop. If you were exercising on a warm day, your rising core body temperature would trigger several homeostatic mechanisms, including increased transport of blood from your core to your body periphery, which is typically cooler. As blood passes through the vessels of the skin, heat would be dissipated to the environment, and the blood returning to your body core would be cooler. In contrast, on a cold day, blood is diverted away from the skin to maintain a warmer body core. In extreme cases, this may result in frostbite.

Blood also helps to maintain the chemical balance of the body. Proteins and other compounds in blood act as buffers, which thereby help to regulate the

pH of body tissues. Blood also helps to regulate the water content of body cells.

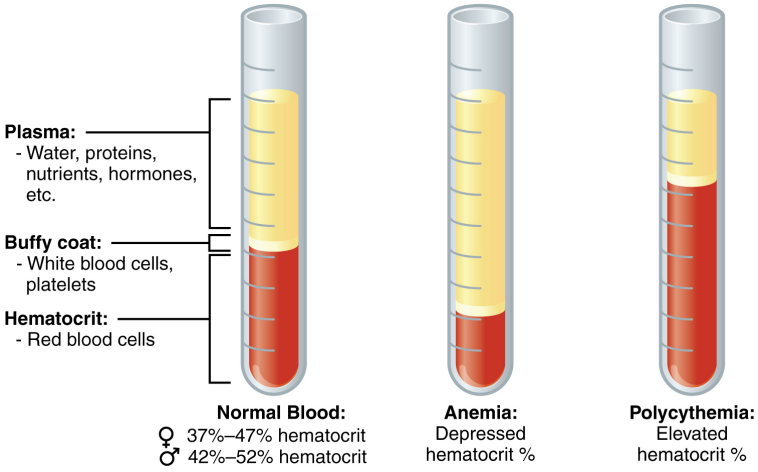
Composition of Blood

You have probably had blood drawn from a superficial vein in your arm, which was then sent to a lab for analysis. Some of the most common blood tests—for instance, those measuring lipid or glucose levels in plasma—determine which substances are present within blood and in what quantities. Other blood tests check for the composition of the blood itself, including the quantities and types of formed elements.

One such test, called a **hematocrit**, measures the percentage of RBCs, clinically known as erythrocytes, in a blood sample. It is performed by spinning the blood sample in a specialized centrifuge, a process that causes the heavier elements suspended within the blood sample to separate from the lightweight, liquid plasma ([link]). Because the heaviest elements in blood are the erythrocytes, these settle at the very bottom of the hematocrit tube. Located above the erythrocytes is a pale, thin layer composed of the remaining formed elements of blood. These are the WBCs, clinically known as leukocytes, and the platelets, cell fragments also called thrombocytes. This layer is referred to as the **buffy coat** because of its color; it normally constitutes less than 1 percent of a blood sample. Above the buffy coat is the blood plasma, normally a pale, straw-colored fluid, which constitutes the remainder of the sample.

The volume of erythrocytes after centrifugation is also commonly referred to as **packed cell volume (PCV)**. In normal blood, about 45 percent of a sample is erythrocytes. The hematocrit of any one sample can vary significantly, however, about 36–50 percent, according to gender and other factors. Normal hematocrit values for females range from 37 to 47, with a mean value of 41; for males, hematocrit ranges from 42 to 52, with a mean of 47. The percentage of other formed elements, the WBCs and platelets, is extremely small so it is not normally considered with the hematocrit. So the mean plasma percentage is the percent of blood that is not erythrocytes: for females, it is approximately 59 (or 100 minus 41), and for males, it is approximately 53 (or 100 minus 47).

Composition of Blood



The cellular elements of blood include a vast number of erythrocytes and comparatively fewer leukocytes and platelets. Plasma is the fluid in which the formed elements are suspended. A sample of blood spun in a centrifuge reveals that plasma is the lightest component. It floats at the top of the tube separated from the heaviest elements, the erythrocytes, by a buffy coat of leukocytes and platelets. Hematocrit is the percentage of the total sample that is comprised of erythrocytes. Depressed and elevated hematocrit levels are shown for comparison.

Characteristics of Blood

When you think about blood, the first characteristic that probably comes to mind is its color. Blood that has just taken up oxygen in the lungs is bright red, and blood that has released oxygen in the tissues is a more dusky red. This is because hemoglobin is a pigment that changes color, depending upon the degree of oxygen saturation.

Blood is viscous and somewhat sticky to the touch. It has a viscosity approximately five times greater than water. Viscosity is a measure of a fluid's thickness or resistance to flow, and is influenced by the presence of the plasma proteins and formed elements within the blood. The viscosity of blood has a dramatic impact on blood pressure and flow. Consider the difference in flow between water and honey. The more viscous honey would demonstrate a greater resistance to flow than the less viscous water. The same principle applies to blood.

The normal temperature of blood is slightly higher than normal body temperature—about 38 °C (or 100.4 °F), compared to 37 °C (or 98.6 °F) for an internal body temperature reading, although daily variations of 0.5 °C are normal. Although the surface of blood vessels is relatively smooth, as blood flows through them, it experiences some friction and resistance, especially as vessels age and lose their elasticity, thereby producing heat. This accounts for its slightly higher temperature.

The pH of blood averages about 7.4; however, it can range from 7.35 to 7.45 in a healthy person. Blood is therefore somewhat more basic (alkaline) on a chemical scale than pure water, which has a pH of 7.0. Blood contains numerous buffers that actually help to regulate pH.

Blood constitutes approximately 8 percent of adult body weight. Adult males typically average about 5 to 6 liters of blood. Females average 4–5 liters.

Blood Plasma

Like other fluids in the body, plasma is composed primarily of water: In fact, it is about 92 percent water. Dissolved or suspended within this water is a mixture of substances, most of which are proteins. There are literally hundreds of substances dissolved or suspended in the plasma, although many of them are found only in very small quantities.

Note:

Visit this site for a list of normal levels established for many of the substances found in a sample of blood. Serum, one of the specimen types included, refers to a sample of plasma after clotting factors have been removed. What types of measurements are given for levels of glucose in the blood?

Plasma Proteins

About 7 percent of the volume of plasma—nearly all that is not water—is made of proteins. These include several plasma proteins (proteins that are unique to the plasma), plus a much smaller number of regulatory proteins, including enzymes and some hormones. The major components of plasma are summarized in [link].

The three major groups of plasma proteins are as follows:

- **Albumin** is the most abundant of the plasma proteins. Manufactured by the liver, albumin molecules serve as binding proteins—transport vehicles for fatty acids and steroid hormones. Recall that lipids are hydrophobic; however, their binding to albumin enables their transport in the watery plasma. Albumin is also the most significant contributor to the osmotic pressure of blood; that is, its presence holds water inside the blood vessels and draws water from the tissues, across blood vessel walls, and into the bloodstream. This in turn helps to maintain both blood volume and blood pressure. Albumin normally accounts for approximately 54 percent of the total plasma protein content, in clinical levels of 3.5–5.0 g/dL blood.
- The second most common plasma proteins are the **globulins**. A heterogeneous group, there are three main subgroups known as alpha, beta, and gamma globulins. The alpha and beta globulins transport

iron, lipids, and the fat-soluble vitamins A, D, E, and K to the cells; like albumin, they also contribute to osmotic pressure. The gamma globulins are proteins involved in immunity and are better known as an **antibodies** or **immunoglobulins**. Although other plasma proteins are produced by the liver, immunoglobulins are produced by specialized leukocytes known as plasma cells. (Seek additional content for more information about immunoglobulins.) Globulins make up approximately 38 percent of the total plasma protein volume, in clinical levels of 1.0–1.5 g/dL blood.

- The least abundant plasma protein is **fibrinogen**. Like albumin and the alpha and beta globulins, fibrinogen is produced by the liver. It is essential for blood clotting, a process described later in this chapter. Fibrinogen accounts for about 7 percent of the total plasma protein volume, in clinical levels of 0.2–0.45 g/dL blood.

Other Plasma Solutes

In addition to proteins, plasma contains a wide variety of other substances. These include various electrolytes, such as sodium, potassium, and calcium ions; dissolved gases, such as oxygen, carbon dioxide, and nitrogen; various organic nutrients, such as vitamins, lipids, glucose, and amino acids; and metabolic wastes. All of these nonprotein solutes combined contribute approximately 1 percent to the total volume of plasma.

Major Blood Components

Component and % of blood	Subcomponent and % of component	Type and % (where appropriate)	Site of production	Major function(s)
Plasma 46–63 percent	Water 92 percent	Fluid	Absorbed by intestinal tract or produced by metabolism	Transport medium
	Plasma proteins 7 percent	Albumin 54–60 percent	Liver	Maintain osmotic concentration, transport lipid molecules
		Globulins 35–38 percent	Alpha globulins—liver	Transport, maintain osmotic concentration
			Beta globulins—liver	Transport, maintain osmotic concentration
			Gamma globulins (immunoglobulins)—plasma cells	Immune responses
		Fibrinogen 4–7 percent	Liver	Blood clotting in hemostasis
	Regulatory proteins <1 percent	Hormones and enzymes	Various sources	Regulate various body functions
	Other solutes 1 percent	Nutrients, gases, and wastes	Absorbed by intestinal tract, exchanged in respiratory system, or produced by cells	Numerous and varied
Formed elements 37–54 percent	Erythrocytes 99 percent	Erythrocytes	Red bone marrow	Transport gases, primarily oxygen and some carbon dioxide
	Leukocytes <1 percent Platelets <1 percent	Granular leukocytes: neutrophils eosinophils basophils	Red bone marrow	Nonspecific immunity
		Agranular leukocytes: lymphocytes monocytes	Lymphocytes: bone marrow and lymphatic tissue	Lymphocytes: specific immunity
			Monocytes: red bone marrow	Monocytes: nonspecific immunity
	Platelets <1 percent		Megakaryocytes: red bone marrow	Hemostasis

Note:

Career Connection

Phlebotomy and Medical Lab Technology

Phlebotomists are professionals trained to draw blood (phleb- = “a blood vessel”; -tomy = “to cut”). When more than a few drops of blood are

required, phlebotomists perform a venipuncture, typically of a surface vein in the arm. They perform a capillary stick on a finger, an earlobe, or the heel of an infant when only a small quantity of blood is required. An arterial stick is collected from an artery and used to analyze blood gases. After collection, the blood may be analyzed by medical laboratories or perhaps used for transfusions, donations, or research. While many allied health professionals practice phlebotomy, the American Society of Phlebotomy Technicians issues certificates to individuals passing a national examination, and some large labs and hospitals hire individuals expressly for their skill in phlebotomy.

Medical or clinical laboratories employ a variety of individuals in technical positions:

- Medical technologists (MT), also known as clinical laboratory technologists (CLT), typically hold a bachelor's degree and certification from an accredited training program. They perform a wide variety of tests on various body fluids, including blood. The information they provide is essential to the primary care providers in determining a diagnosis and in monitoring the course of a disease and response to treatment.
- Medical laboratory technicians (MLT) typically have an associate's degree but may perform duties similar to those of an MT.
- Medical laboratory assistants (MLA) spend the majority of their time processing samples and carrying out routine assignments within the lab. Clinical training is required, but a degree may not be essential to obtaining a position.

Chapter Review

Blood is a fluid connective tissue critical to the transportation of nutrients, gases, and wastes throughout the body; to defend the body against infection and other threats; and to the homeostatic regulation of pH, temperature, and other internal conditions. Blood is composed of formed elements—erythrocytes, leukocytes, and cell fragments called platelets—and a fluid extracellular matrix called plasma. More than 90 percent of plasma is water.

The remainder is mostly plasma proteins—mainly albumin, globulins, and fibrinogen—and other dissolved solutes such as glucose, lipids, electrolytes, and dissolved gases. Because of the formed elements and the plasma proteins and other solutes, blood is sticky and more viscous than water. It is also slightly alkaline, and its temperature is slightly higher than normal body temperature.

Interactive Link Questions

Exercise:

Problem:

Visit this site for a list of normal levels established for many of the substances found in a sample of blood. Serum, one of the specimen types included, refers to a sample of plasma after clotting factors have been removed. What types of measurements are given for levels of glucose in the blood?

Solution:

There are values given for percent saturation, tension, and blood gas, and there are listings for different types of hemoglobin.

Review Questions

Exercise:

Problem: Which of the following statements about blood is true?

a. Blood is about 92 percent water.
b. Blood is slightly more acidic than water.
c. Blood is slightly more viscous than water.
d. Blood is slightly more salty than seawater.

Solution:

C

Exercise:

Problem: Which of the following statements about albumin is true?

a. It draws water out of the blood vessels and into the body’s tissues.
b. It is the most abundant plasma protein.
c. It is produced by specialized leukocytes called plasma cells.
d. All of the above are true.

Solution:

B

Exercise:

Problem:

Which of the following plasma proteins is *not* produced by the liver?

a. fibrinogen
b. alpha globulin
c. beta globulin
d. immunoglobulin

Solution:

D

Critical Thinking Questions

Exercise:

Problem:

A patient’s hematocrit is 42 percent. Approximately what percentage of the patient’s blood is plasma?

Solution:

The patient’s blood is approximately 58 percent plasma (since the buffy coat is less than 1 percent).

Exercise:

Problem:

Why would it be incorrect to refer to the formed elements as cells?

Solution:

The formed elements include erythrocytes and leukocytes, which are cells (although mature erythrocytes do not have a nucleus); however, the formed elements also include platelets, which are not true cells but cell fragments.

Exercise:

Problem:

True or false: The buffy coat is the portion of a blood sample that is made up of its proteins.

Solution:

False. The buffy coat is the portion of blood that is made up of its leukocytes and platelets.

Glossary

albumin

most abundant plasma protein, accounting for most of the osmotic pressure of plasma

antibodies
: (also, immunoglobulins or gamma globulins) antigen-specific proteins produced by specialized B lymphocytes that protect the body by binding to foreign objects such as bacteria and viruses

blood
: liquid connective tissue composed of formed elements—erythrocytes, leukocytes, and platelets—and a fluid extracellular matrix called plasma; component of the cardiovascular system

buffy coat
: thin, pale layer of leukocytes and platelets that separates the erythrocytes from the plasma in a sample of centrifuged blood

fibrinogen
: plasma protein produced in the liver and involved in blood clotting

formed elements
: cellular components of blood; that is, erythrocytes, leukocytes, and platelets

globulins
: heterogeneous group of plasma proteins that includes transport proteins, clotting factors, immune proteins, and others

hematocrit
: (also, packed cell volume) volume percentage of erythrocytes in a sample of centrifuged blood

immunoglobulins
: (also, antibodies or gamma globulins) antigen-specific proteins produced by specialized B lymphocytes that protect the body by binding to foreign objects such as bacteria and viruses

packed cell volume (PCV)

(also, hematocrit) volume percentage of erythrocytes present in a sample of centrifuged blood

plasma
: in blood, the liquid extracellular matrix composed mostly of water that circulates the formed elements and dissolved materials throughout the cardiovascular system

platelets
: (also, thrombocytes) one of the formed elements of blood that consists of cell fragments broken off from megakaryocytes

red blood cells (RBCs)
: (also, erythrocytes) one of the formed elements of blood that transports oxygen

white blood cells (WBCs)
: (also, leukocytes) one of the formed elements of blood that provides defense against disease agents and foreign materials

Blood Flow, Blood Pressure, and Resistance
By the end of this section, you will be able to:

- Distinguish between systolic pressure, diastolic pressure, pulse pressure, and mean arterial pressure
- Describe the clinical measurement of pulse and blood pressure
- Identify and discuss five variables affecting arterial blood flow and blood pressure
- Discuss several factors affecting blood flow in the venous system

Blood flow refers to the movement of blood through a vessel, tissue, or organ, and is usually expressed in terms of volume of blood per unit of time. It is initiated by the contraction of the ventricles of the heart. Ventricular contraction ejects blood into the major arteries, resulting in flow from regions of higher pressure to regions of lower pressure, as blood encounters smaller arteries and arterioles, then capillaries, then the venules and veins of the venous system. This section discusses a number of critical variables that contribute to blood flow throughout the body. It also discusses the factors that impede or slow blood flow, a phenomenon known as **resistance**.

As noted earlier, hydrostatic pressure is the force exerted by a fluid due to gravitational pull, usually against the wall of the container in which it is located. One form of hydrostatic pressure is **blood pressure**, the force exerted by blood upon the walls of the blood vessels or the chambers of the heart. Blood pressure may be measured in capillaries and veins, as well as the vessels of the pulmonary circulation; however, the term blood pressure without any specific descriptors typically refers to systemic arterial blood pressure—that is, the pressure of blood flowing in the arteries of the systemic circulation. In clinical practice, this pressure is measured in mm Hg and is usually obtained using the brachial artery of the arm.

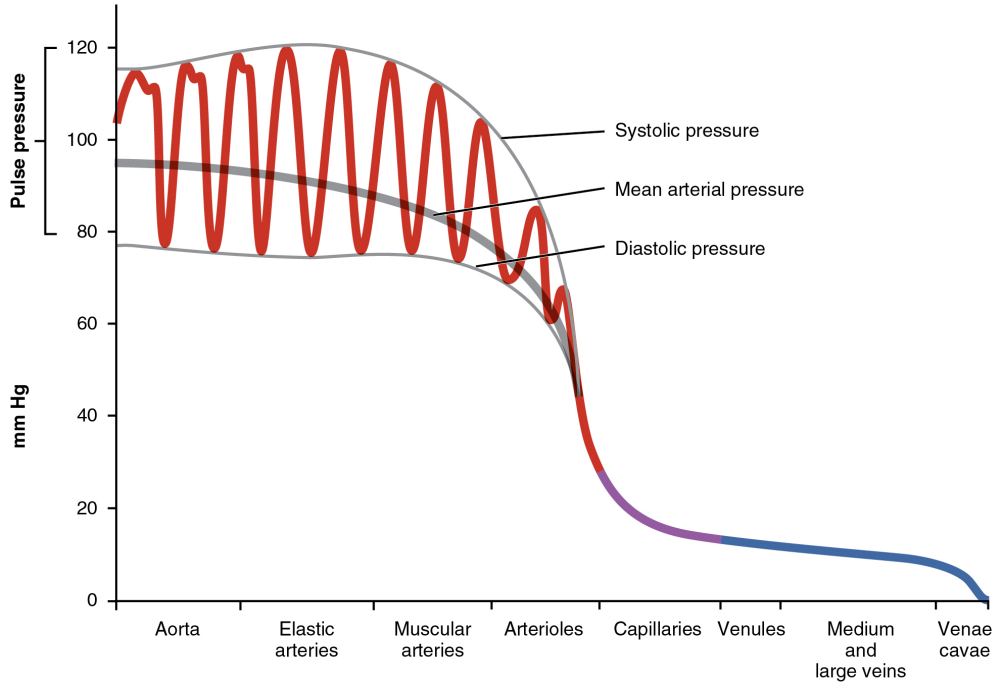
Components of Arterial Blood Pressure

Arterial blood pressure in the larger vessels consists of several distinct components ([link]): systolic and diastolic pressures, pulse pressure, and mean arterial pressure.

Systolic and Diastolic Pressures

When systemic arterial blood pressure is measured, it is recorded as a ratio of two numbers (e.g., 120/80 is a normal adult blood pressure), expressed as systolic pressure over diastolic pressure. The **systolic pressure** is the higher value (typically around 120 mm Hg) and reflects the arterial pressure resulting from the ejection of blood during ventricular contraction, or systole. The **diastolic pressure** is the lower value (usually about 80 mm Hg) and represents the arterial pressure of blood during ventricular relaxation, or diastole.

Systemic Blood Pressure



The graph shows the components of blood pressure throughout the blood vessels, including systolic, diastolic, mean arterial, and pulse pressures.

Pulse Pressure

As shown in [link], the difference between the systolic pressure and the diastolic pressure is the **pulse pressure**. For example, an individual with a systolic pressure of 120 mm Hg and a diastolic pressure of 80 mm Hg would have a pulse pressure of 40 mmHg.

Generally, a pulse pressure should be at least 25 percent of the systolic pressure. A pulse pressure below this level is described as low or narrow. This may occur, for example, in patients with a low stroke volume, which may be seen in congestive heart failure, stenosis of the aortic valve, or significant blood loss following trauma. In contrast, a high or wide pulse pressure is common in healthy people following strenuous exercise, when their resting pulse pressure of 30–40 mm Hg may increase temporarily to 100 mm Hg as stroke volume increases. A persistently high pulse pressure at or above 100 mm Hg may indicate excessive resistance in the arteries and can be caused by a variety of disorders. Chronic high resting pulse pressures can degrade the heart, brain, and kidneys, and warrant medical treatment.

Mean Arterial Pressure

Mean arterial pressure (MAP) represents the “average” pressure of blood in the arteries, that is, the average force driving blood into vessels that serve the tissues. Mean is a statistical concept and is calculated by taking the sum of the values divided by the number of values. Although complicated to measure directly and complicated to calculate, MAP can be approximated by adding the diastolic pressure to one-third of the pulse pressure or systolic pressure minus the diastolic pressure:

$$\text{MAP} = \text{diastolic BP} + \frac{(\text{systolic-diastolic BP})}{3}$$

In [link], this value is approximately 80 + (120 − 80) / 3, or 93.33. Normally, the MAP falls within the range of 70–110 mm Hg. If the value falls below 60 mm Hg for an extended time, blood pressure will not be high enough to ensure circulation to and through the tissues, which results in **ischemia**, or insufficient blood flow. A condition called **hypoxia**,

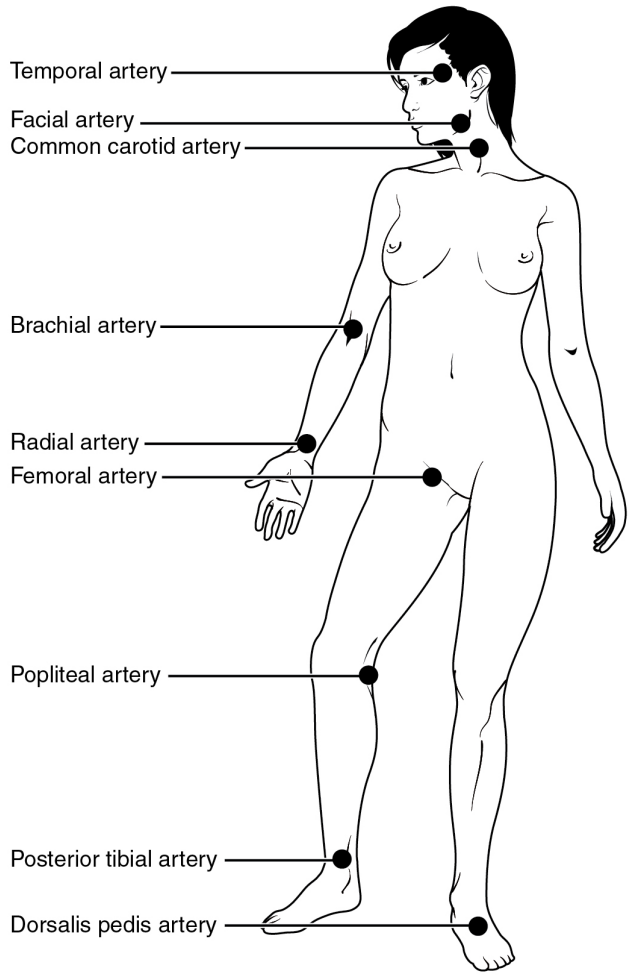
inadequate oxygenation of tissues, commonly accompanies ischemia. The term hypoxemia refers to low levels of oxygen in systemic arterial blood. Neurons are especially sensitive to hypoxia and may die or be damaged if blood flow and oxygen supplies are not quickly restored.

Pulse

After blood is ejected from the heart, elastic fibers in the arteries help maintain a high-pressure gradient as they expand to accommodate the blood, then recoil. This expansion and recoiling effect, known as the **pulse**, can be palpated manually or measured electronically. Although the effect diminishes over distance from the heart, elements of the systolic and diastolic components of the pulse are still evident down to the level of the arterioles.

Because pulse indicates heart rate, it is measured clinically to provide clues to a patient's state of health. It is recorded as beats per minute. Both the rate and the strength of the pulse are important clinically. A high or irregular pulse rate can be caused by physical activity or other temporary factors, but it may also indicate a heart condition. The pulse strength indicates the strength of ventricular contraction and cardiac output. If the pulse is strong, then systolic pressure is high. If it is weak, systolic pressure has fallen, and medical intervention may be warranted.

Pulse can be palpated manually by placing the tips of the fingers across an artery that runs close to the body surface and pressing lightly. While this procedure is normally performed using the radial artery in the wrist or the common carotid artery in the neck, any superficial artery that can be palpated may be used ([link]). Common sites to find a pulse include temporal and facial arteries in the head, brachial arteries in the upper arm, femoral arteries in the thigh, popliteal arteries behind the knees, posterior tibial arteries near the medial tarsal regions, and dorsalis pedis arteries in the feet. A variety of commercial electronic devices are also available to measure pulse.
Pulse Sites



The pulse is most readily measured at the radial artery, but can be measured at any of the pulse points shown.

Measurement of Blood Pressure

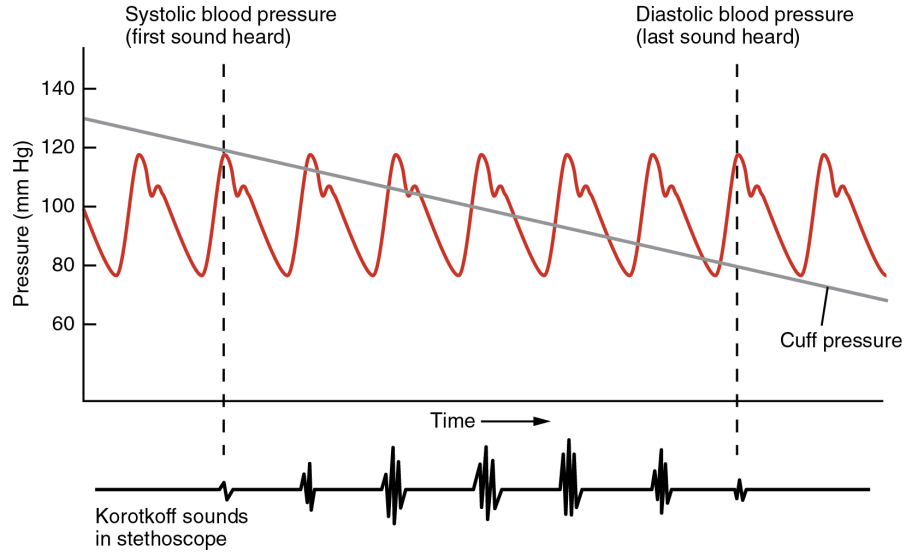
Blood pressure is one of the critical parameters measured on virtually every patient in every healthcare setting. The technique used today was developed more than 100 years ago by a pioneering Russian physician, Dr. Nikolai Korotkoff. Turbulent blood flow through the vessels can be heard as a soft ticking while measuring blood pressure; these sounds are known as

Korotkoff sounds. The technique of measuring blood pressure requires the use of a **sphygmomanometer** (a blood pressure cuff attached to a measuring device) and a stethoscope. The technique is as follows:

- The clinician wraps an inflatable cuff tightly around the patient's arm at about the level of the heart.
- The clinician squeezes a rubber pump to inject air into the cuff, raising pressure around the artery and temporarily cutting off blood flow into the patient's arm.
- The clinician places the stethoscope on the patient's antecubital region and, while gradually allowing air within the cuff to escape, listens for the Korotkoff sounds.

Although there are five recognized Korotkoff sounds, only two are normally recorded. Initially, no sounds are heard since there is no blood flow through the vessels, but as air pressure drops, the cuff relaxes, and blood flow returns to the arm. As shown in [link], the first sound heard through the stethoscope—the first Korotkoff sound—indicates systolic pressure. As more air is released from the cuff, blood is able to flow freely through the brachial artery and all sounds disappear. The point at which the last sound is heard is recorded as the patient's diastolic pressure.

Blood Pressure Measurement



When pressure in a sphygmomanometer cuff is released, a clinician can hear the Korotkoff sounds.

In this graph, a blood pressure tracing is aligned to a measurement of systolic and diastolic pressures.

The majority of hospitals and clinics have automated equipment for measuring blood pressure that work on the same principles. An even more recent innovation is a small instrument that wraps around a patient's wrist. The patient then holds the wrist over the heart while the device measures blood flow and records pressure.

Variables Affecting Blood Flow and Blood Pressure

Five variables influence blood flow and blood pressure:

- Cardiac output
- Compliance
- Volume of the blood
- Viscosity of the blood
- Blood vessel length and diameter

Recall that blood moves from higher pressure to lower pressure. It is pumped from the heart into the arteries at high pressure. If you increase pressure in the arteries (afterload), and cardiac function does not compensate, blood flow will actually decrease. In the venous system, the opposite relationship is true. Increased pressure in the veins does not decrease flow as it does in arteries, but actually increases flow. Since pressure in the veins is normally relatively low, for blood to flow back into the heart, the pressure in the atria during atrial diastole must be even lower. It normally approaches zero, except when the atria contract (see [link]).

Cardiac Output

Cardiac output is the measurement of blood flow from the heart through the ventricles, and is usually measured in liters per minute. Any factor that

causes cardiac output to increase, by elevating heart rate or stroke volume or both, will elevate blood pressure and promote blood flow. These factors include sympathetic stimulation, the catecholamines epinephrine and norepinephrine, thyroid hormones, and increased calcium ion levels. Conversely, any factor that decreases cardiac output, by decreasing heart rate or stroke volume or both, will decrease arterial pressure and blood flow. These factors include parasympathetic stimulation, elevated or decreased potassium ion levels, decreased calcium levels, anoxia, and acidosis.

Compliance

Compliance is the ability of any compartment to expand to accommodate increased content. A metal pipe, for example, is not compliant, whereas a balloon is. The greater the compliance of an artery, the more effectively it is able to expand to accommodate surges in blood flow without increased resistance or blood pressure. Veins are more compliant than arteries and can expand to hold more blood. When vascular disease causes stiffening of arteries, compliance is reduced and resistance to blood flow is increased. The result is more turbulence, higher pressure within the vessel, and reduced blood flow. This increases the work of the heart.

A Mathematical Approach to Factors Affecting Blood Flow

Jean Louis Marie Poiseuille was a French physician and physiologist who devised a mathematical equation describing blood flow and its relationship to known parameters. The same equation also applies to engineering studies of the flow of fluids. Although understanding the math behind the relationships among the factors affecting blood flow is not necessary to understand blood flow, it can help solidify an understanding of their relationships. Please note that even if the equation looks intimidating, breaking it down into its components and following the relationships will make these relationships clearer, even if you are weak in math. Focus on the three critical variables: radius (r), vessel length (λ), and viscosity (η).

Poiseuille’s equation:
Equation:

$$\text{Blood flow} = \frac{\pi\ \Delta P\ r^4}{8\eta\lambda}$$

π is the Greek letter pi, used to represent the mathematical constant that is the ratio of a circle’s circumference to its diameter. It may commonly be represented as 3.14, although the actual number extends to infinity.
ΔP represents the difference in pressure.
r^4 is the radius (one-half of the diameter) of the vessel to the fourth power.
η is the Greek letter eta and represents the viscosity of the blood.
λ is the Greek letter lambda and represents the length of a blood vessel.

One of several things this equation allows us to do is calculate the resistance in the vascular system. Normally this value is extremely difficult to measure, but it can be calculated from this known relationship:
Equation:

$$\text{Blood flow} = \frac{\Delta P}{\text{Resistance}}$$

If we rearrange this slightly,
Equation:

$$\text{Resistance} = \frac{\Delta P}{\text{Blood flow}}$$

Then by substituting Pouseille’s equation for blood flow:
Equation:

$$\text{Resistance} = \frac{8\eta\lambda}{\pi r^4}$$

By examining this equation, you can see that there are only three variables: viscosity, vessel length, and radius, since 8 and π are both constants. The important thing to remember is this: Two of these variables, viscosity and vessel length, will change slowly in the body. Only one of these factors, the radius, can be changed rapidly by vasoconstriction and vasodilation, thus dramatically impacting resistance and flow. Further, small changes in the radius will greatly affect flow, since it is raised to the fourth power in the equation.

We have briefly considered how cardiac output and blood volume impact blood flow and pressure; the next step is to see how the other variables (contraction, vessel length, and viscosity) articulate with Pouseille's equation and what they can teach us about the impact on blood flow.

Blood Volume

The relationship between blood volume, blood pressure, and blood flow is intuitively obvious. Water may merely trickle along a creek bed in a dry season, but rush quickly and under great pressure after a heavy rain. Similarly, as blood volume decreases, pressure and flow decrease. As blood volume increases, pressure and flow increase.

Under normal circumstances, blood volume varies little. Low blood volume, called **hypovolemia**, may be caused by bleeding, dehydration, vomiting, severe burns, or some medications used to treat hypertension. It is important to recognize that other regulatory mechanisms in the body are so effective at maintaining blood pressure that an individual may be asymptomatic until 10–20 percent of the blood volume has been lost. Treatment typically includes intravenous fluid replacement.

Hypervolemia, excessive fluid volume, may be caused by retention of water and sodium, as seen in patients with heart failure, liver cirrhosis, some forms of kidney disease, hyperaldosteronism, and some glucocorticoid steroid treatments. Restoring homeostasis in these patients depends upon reversing the condition that triggered the hypervolemia.

Blood Viscosity

Viscosity is the thickness of fluids that affects their ability to flow. Clean water, for example, is less viscous than mud. The viscosity of blood is directly proportional to resistance and inversely proportional to flow; therefore, any condition that causes viscosity to increase will also increase resistance and decrease flow. For example, imagine sipping milk, then a milkshake, through the same size straw. You experience more resistance and therefore less flow from the milkshake. Conversely, any condition that causes viscosity to decrease (such as when the milkshake melts) will decrease resistance and increase flow.

Normally the viscosity of blood does not change over short periods of time. The two primary determinants of blood viscosity are the formed elements and plasma proteins. Since the vast majority of formed elements are erythrocytes, any condition affecting erythropoiesis, such as polycythemia or anemia, can alter viscosity. Since most plasma proteins are produced by the liver, any condition affecting liver function can also change the viscosity slightly and therefore alter blood flow. Liver abnormalities such as hepatitis, cirrhosis, alcohol damage, and drug toxicities result in decreased levels of plasma proteins, which decrease blood viscosity. While leukocytes and platelets are normally a small component of the formed elements, there are some rare conditions in which severe overproduction can impact viscosity as well.

Vessel Length and Diameter

The length of a vessel is directly proportional to its resistance: the longer the vessel, the greater the resistance and the lower the flow. As with blood volume, this makes intuitive sense, since the increased surface area of the vessel will impede the flow of blood. Likewise, if the vessel is shortened, the resistance will decrease and flow will increase.

The length of our blood vessels increases throughout childhood as we grow, of course, but is unchanging in adults under normal physiological circumstances. Further, the distribution of vessels is not the same in all

tissues. Adipose tissue does not have an extensive vascular supply. One pound of adipose tissue contains approximately 200 miles of vessels, whereas skeletal muscle contains more than twice that. Overall, vessels decrease in length only during loss of mass or amputation. An individual weighing 150 pounds has approximately 60,000 miles of vessels in the body. Gaining about 10 pounds adds from 2000 to 4000 miles of vessels, depending upon the nature of the gained tissue. One of the great benefits of weight reduction is the reduced stress to the heart, which does not have to overcome the resistance of as many miles of vessels.

In contrast to length, the diameter of blood vessels changes throughout the body, according to the type of vessel, as we discussed earlier. The diameter of any given vessel may also change frequently throughout the day in response to neural and chemical signals that trigger vasodilation and vasoconstriction. The **vascular tone** of the vessel is the contractile state of the smooth muscle and the primary determinant of diameter, and thus of resistance and flow. The effect of vessel diameter on resistance is inverse: Given the same volume of blood, an increased diameter means there is less blood contacting the vessel wall, thus lower friction and lower resistance, subsequently increasing flow. A decreased diameter means more of the blood contacts the vessel wall, and resistance increases, subsequently decreasing flow.

The influence of lumen diameter on resistance is dramatic: A slight increase or decrease in diameter causes a huge decrease or increase in resistance. This is because resistance is inversely proportional to the radius of the blood vessel (one-half of the vessel's diameter) raised to the fourth power ($R = 1/r^4$). This means, for example, that if an artery or arteriole constricts to one-half of its original radius, the resistance to flow will increase 16 times. And if an artery or arteriole dilates to twice its initial radius, then resistance in the vessel will decrease to 1/16 of its original value and flow will increase 16 times.

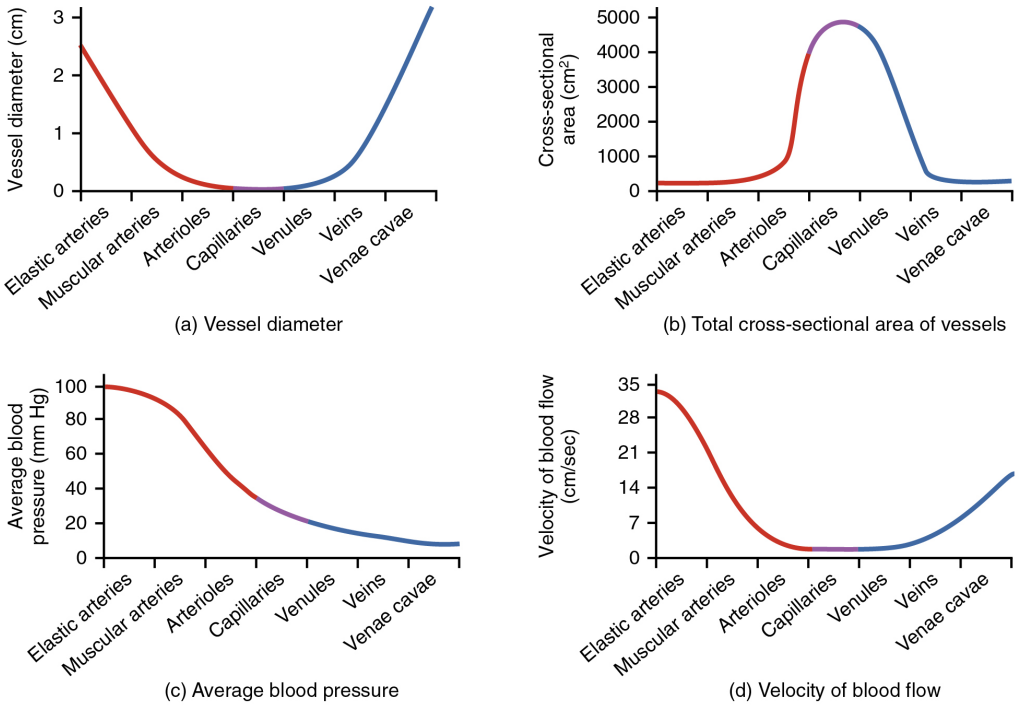
The Roles of Vessel Diameter and Total Area in Blood Flow and Blood Pressure

Recall that we classified arterioles as resistance vessels, because given their small lumen, they dramatically slow the flow of blood from arteries. In fact, arterioles are the site of greatest resistance in the entire vascular network. This may seem surprising, given that capillaries have a smaller size. How can this phenomenon be explained?

[link] compares vessel diameter, total cross-sectional area, average blood pressure, and blood velocity through the systemic vessels. Notice in parts (a) and (b) that the total cross-sectional area of the body's capillary beds is far greater than any other type of vessel. Although the diameter of an individual capillary is significantly smaller than the diameter of an arteriole, there are vastly more capillaries in the body than there are other types of blood vessels. Part (c) shows that blood pressure drops unevenly as blood travels from arteries to arterioles, capillaries, venules, and veins, and encounters greater resistance. However, the site of the most precipitous drop, and the site of greatest resistance, is the arterioles. This explains why vasodilation and vasoconstriction of arterioles play more significant roles in regulating blood pressure than do the vasodilation and vasoconstriction of other vessels.

Part (d) shows that the velocity (speed) of blood flow decreases dramatically as the blood moves from arteries to arterioles to capillaries. This slow flow rate allows more time for exchange processes to occur. As blood flows through the veins, the rate of velocity increases, as blood is returned to the heart.

Relationships among Vessels in the Systemic Circuit



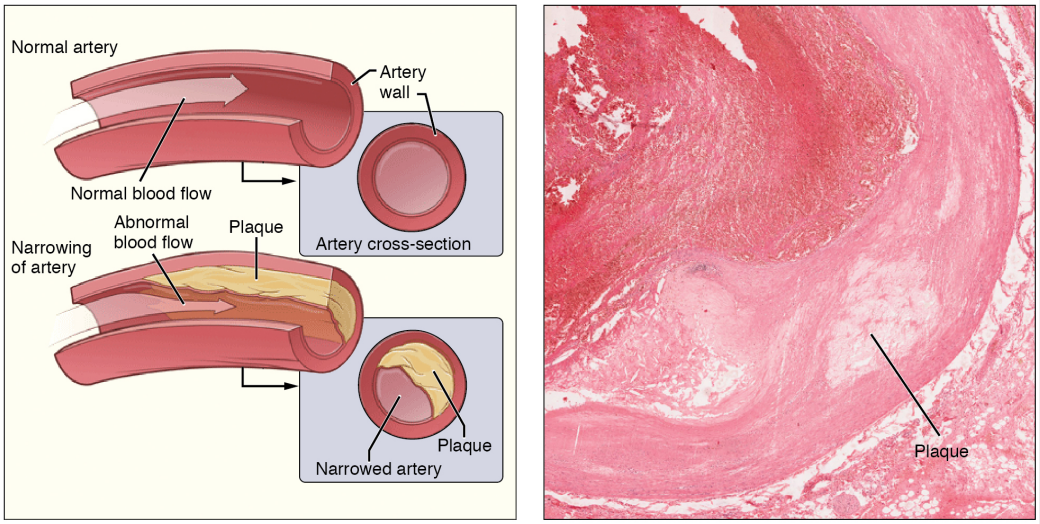
The relationships among blood vessels that can be compared include (a) vessel diameter, (b) total cross-sectional area, (c) average blood pressure, and (d) velocity of blood flow.

Note:
Disorders of the…
Cardiovascular System: Arteriosclerosis
Compliance allows an artery to expand when blood is pumped through it from the heart, and then to recoil after the surge has passed. This helps promote blood flow. In arteriosclerosis, compliance is reduced, and pressure and resistance within the vessel increase. This is a leading cause of hypertension and coronary heart disease, as it causes the heart to work harder to generate a pressure great enough to overcome the resistance. Arteriosclerosis begins with injury to the endothelium of an artery, which may be caused by irritation from high blood glucose, infection, tobacco use, excessive blood lipids, and other factors. Artery walls that are

constantly stressed by blood flowing at high pressure are also more likely to be injured—which means that hypertension can promote arteriosclerosis, as well as result from it.

Recall that tissue injury causes inflammation. As inflammation spreads into the artery wall, it weakens and scars it, leaving it stiff (sclerotic). As a result, compliance is reduced. Moreover, circulating triglycerides and cholesterol can seep between the damaged lining cells and become trapped within the artery wall, where they are frequently joined by leukocytes, calcium, and cellular debris. Eventually, this buildup, called plaque, can narrow arteries enough to impair blood flow. The term for this condition, atherosclerosis (athero- = “porridge”) describes the mealy deposits ([link]).

Atherosclerosis



(a) Atherosclerosis can result from plaques formed by the buildup of fatty, calcified deposits in an artery. (b) Plaques can also take other forms, as shown in this micrograph of a coronary artery that has a buildup of connective tissue within the artery wall. LM × 40. (Micrograph provided by the Regents of University of Michigan Medical School © 2012)

Sometimes a plaque can rupture, causing microscopic tears in the artery wall that allow blood to leak into the tissue on the other side. When this happens, platelets rush to the site to clot the blood. This clot can further

obstruct the artery and—if it occurs in a coronary or cerebral artery—cause a sudden heart attack or stroke. Alternatively, plaque can break off and travel through the bloodstream as an embolus until it blocks a more distant, smaller artery.

Even without total blockage, vessel narrowing leads to ischemia—reduced blood flow—to the tissue region “downstream” of the narrowed vessel. Ischemia in turn leads to hypoxia—decreased supply of oxygen to the tissues. Hypoxia involving cardiac muscle or brain tissue can lead to cell death and severe impairment of brain or heart function.

A major risk factor for both arteriosclerosis and atherosclerosis is advanced age, as the conditions tend to progress over time. Arteriosclerosis is normally defined as the more generalized loss of compliance, “hardening of the arteries,” whereas atherosclerosis is a more specific term for the build-up of plaque in the walls of the vessel and is a specific type of arteriosclerosis. There is also a distinct genetic component, and pre-existing hypertension and/or diabetes also greatly increase the risk. However, obesity, poor nutrition, lack of physical activity, and tobacco use all are major risk factors.

Treatment includes lifestyle changes, such as weight loss, smoking cessation, regular exercise, and adoption of a diet low in sodium and saturated fats. Medications to reduce cholesterol and blood pressure may be prescribed. For blocked coronary arteries, surgery is warranted. In angioplasty, a catheter is inserted into the vessel at the point of narrowing, and a second catheter with a balloon-like tip is inflated to widen the opening. To prevent subsequent collapse of the vessel, a small mesh tube called a stent is often inserted. In an endarterectomy, plaque is surgically removed from the walls of a vessel. This operation is typically performed on the carotid arteries of the neck, which are a prime source of oxygenated blood for the brain. In a coronary bypass procedure, a non-vital superficial vessel from another part of the body (often the great saphenous vein) or a synthetic vessel is inserted to create a path around the blocked area of a coronary artery.

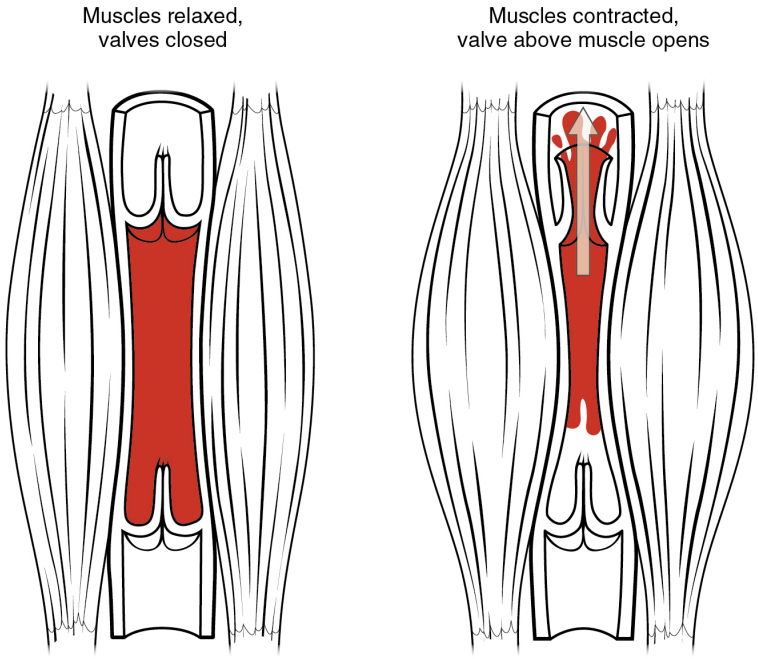
Venous System

The pumping action of the heart propels the blood into the arteries, from an area of higher pressure toward an area of lower pressure. If blood is to flow from the veins back into the heart, the pressure in the veins must be greater than the pressure in the atria of the heart. Two factors help maintain this pressure gradient between the veins and the heart. First, the pressure in the atria during diastole is very low, often approaching zero when the atria are relaxed (atrial diastole). Second, two physiologic “pumps” increase pressure in the venous system. The use of the term “pump” implies a physical device that speeds flow. These physiological pumps are less obvious.

Skeletal Muscle Pump

In many body regions, the pressure within the veins can be increased by the contraction of the surrounding skeletal muscle. This mechanism, known as the **skeletal muscle pump** ([link]), helps the lower-pressure veins counteract the force of gravity, increasing pressure to move blood back to the heart. As leg muscles contract, for example during walking or running, they exert pressure on nearby veins with their numerous one-way valves. This increased pressure causes blood to flow upward, opening valves superior to the contracting muscles so blood flows through. Simultaneously, valves inferior to the contracting muscles close; thus, blood should not seep back downward toward the feet. Military recruits are trained to flex their legs slightly while standing at attention for prolonged periods. Failure to do so may allow blood to pool in the lower limbs rather than returning to the heart. Consequently, the brain will not receive enough oxygenated blood, and the individual may lose consciousness.

Skeletal Muscle Pump



The contraction of skeletal muscles surrounding a vein compresses the blood and increases the pressure in that area. This action forces blood closer to the heart where venous pressure is lower. Note the importance of the one-way valves to assure that blood flows only in the proper direction.

Respiratory Pump

The **respiratory pump** aids blood flow through the veins of the thorax and abdomen. During inhalation, the volume of the thorax increases, largely through the contraction of the diaphragm, which moves downward and compresses the abdominal cavity. The elevation of the chest caused by the contraction of the external intercostal muscles also contributes to the increased volume of the thorax. The volume increase causes air pressure within the thorax to decrease, allowing us to inhale. Additionally, as air

pressure within the thorax drops, blood pressure in the thoracic veins also decreases, falling below the pressure in the abdominal veins. This causes blood to flow along its pressure gradient from veins outside the thorax, where pressure is higher, into the thoracic region, where pressure is now lower. This in turn promotes the return of blood from the thoracic veins to the atria. During exhalation, when air pressure increases within the thoracic cavity, pressure in the thoracic veins increases, speeding blood flow into the heart while valves in the veins prevent blood from flowing backward from the thoracic and abdominal veins.

Pressure Relationships in the Venous System

Although vessel diameter increases from the smaller venules to the larger veins and eventually to the venae cavae (singular = vena cava), the total cross-sectional area actually decreases (see [link]**a** and **b**). The individual veins are larger in diameter than the venules, but their total number is much lower, so their total cross-sectional area is also lower.

Also notice that, as blood moves from venules to veins, the average blood pressure drops (see [link]**c**), but the blood velocity actually increases (see [link]). This pressure gradient drives blood back toward the heart. Again, the presence of one-way valves and the skeletal muscle and respiratory pumps contribute to this increased flow. Since approximately 64 percent of the total blood volume resides in systemic veins, any action that increases the flow of blood through the veins will increase venous return to the heart. Maintaining vascular tone within the veins prevents the veins from merely distending, dampening the flow of blood, and as you will see, vasoconstriction actually enhances the flow.

The Role of Venoconstriction in Resistance, Blood Pressure, and Flow

As previously discussed, vasoconstriction of an artery or arteriole decreases the radius, increasing resistance and pressure, but decreasing flow. Venoconstriction, on the other hand, has a very different outcome. The

walls of veins are thin but irregular; thus, when the smooth muscle in those walls constricts, the lumen becomes more rounded. The more rounded the lumen, the less surface area the blood encounters, and the less resistance the vessel offers. Vasoconstriction increases pressure within a vein as it does in an artery, but in veins, the increased pressure increases flow. Recall that the pressure in the atria, into which the venous blood will flow, is very low, approaching zero for at least part of the relaxation phase of the cardiac cycle. Thus, venoconstriction increases the return of blood to the heart. Another way of stating this is that venoconstriction increases the preload or stretch of the cardiac muscle and increases contraction.

Chapter Review

Blood flow is the movement of blood through a vessel, tissue, or organ. The slowing or blocking of blood flow is called resistance. Blood pressure is the force that blood exerts upon the walls of the blood vessels or chambers of the heart. The components of blood pressure include systolic pressure, which results from ventricular contraction, and diastolic pressure, which results from ventricular relaxation. Pulse pressure is the difference between systolic and diastolic measures, and mean arterial pressure is the “average” pressure of blood in the arterial system, driving blood into the tissues. Pulse, the expansion and recoiling of an artery, reflects the heartbeat. The variables affecting blood flow and blood pressure in the systemic circulation are cardiac output, compliance, blood volume, blood viscosity, and the length and diameter of the blood vessels. In the arterial system, vasodilation and vasoconstriction of the arterioles is a significant factor in systemic blood pressure: Slight vasodilation greatly decreases resistance and increases flow, whereas slight vasoconstriction greatly increases resistance and decreases flow. In the arterial system, as resistance increases, blood pressure increases and flow decreases. In the venous system, constriction increases blood pressure as it does in arteries; the increasing pressure helps to return blood to the heart. In addition, constriction causes the vessel lumen to become more rounded, decreasing resistance and increasing blood flow. Venoconstriction, while less important than arterial vasoconstriction, works with the skeletal muscle pump, the respiratory pump, and their valves to promote venous return to the heart.

Review Questions

Exercise:

Problem:

In a blood pressure measurement of 110/70, the number 70 is the _____.

a. systolic pressure
b. diastolic pressure
c. pulse pressure
d. mean arterial pressure

Solution:

B

Exercise:

Problem: A healthy elastic artery _____.

a. is compliant
b. reduces blood flow
c. is a resistance artery
d. has a thin wall and irregular lumen

Solution:

A

Exercise:

Problem: Which of the following statements is *true*?

a. The longer the vessel, the lower the resistance and the greater the flow.

b. As blood volume decreases, blood pressure and blood flow also decrease.
c. Increased viscosity increases blood flow.
d. All of the above are true.

Solution:

B

Exercise:

Problem: Slight vasodilation in an arteriole prompts a _____.

a. slight increase in resistance
b. huge increase in resistance
c. slight decrease in resistance
d. huge decrease in resistance

Solution:

D

Exercise:

Problem: Venoconstriction increases which of the following?

a. blood pressure within the vein
b. blood flow within the vein
c. return of blood to the heart
d. all of the above

Solution:

D

Critical Thinking Questions

Exercise:

Problem:

You measure a patient’s blood pressure at 130/85. Calculate the patient’s pulse pressure and mean arterial pressure. Determine whether each pressure is low, normal, or high.

Solution:

The patient’s pulse pressure is 130 – 85 = 45 mm Hg. Generally, a pulse pressure should be at least 25 percent of the systolic pressure, but not more than 100 mm Hg. Since 25 percent of 130 = 32.5, the patient’s pulse pressure of 45 is normal. The patient’s mean arterial pressure is 85 + 1/3 (45) = 85 + 15 = 100. Normally, the mean arterial blood pressure falls within the range of 70 – 110 mmHg, so 100 is normal.

Exercise:

Problem:

An obese patient comes to the clinic complaining of swollen feet and ankles, fatigue, shortness of breath, and often feeling “spaced out.” She is a cashier in a grocery store, a job that requires her to stand all day. Outside of work, she engages in no physical activity. She confesses that, because of her weight, she finds even walking uncomfortable. Explain how the skeletal muscle pump might play a role in this patient’s signs and symptoms.

Solution:

People who stand upright all day and are inactive overall have very little skeletal muscle activity in the legs. Pooling of blood in the legs and feet is common. Venous return to the heart is reduced, a condition that in turn reduces cardiac output and therefore oxygenation of tissues throughout the body. This could at least partially account for the

patient’s fatigue and shortness of breath, as well as her “spaced out” feeling, which commonly reflects reduced oxygen to the brain.

Glossary

blood flow
: movement of blood through a vessel, tissue, or organ that is usually expressed in terms of volume per unit of time

blood pressure
: force exerted by the blood against the wall of a vessel or heart chamber; can be described with the more generic term hydrostatic pressure

compliance
: degree to which a blood vessel can stretch as opposed to being rigid

diastolic pressure
: lower number recorded when measuring arterial blood pressure; represents the minimal value corresponding to the pressure that remains during ventricular relaxation

hypervolemia
: abnormally high levels of fluid and blood within the body

hypovolemia
: abnormally low levels of fluid and blood within the body

hypoxia
: lack of oxygen supply to the tissues

ischemia
: insufficient blood flow to the tissues

Korotkoff sounds
: noises created by turbulent blood flow through the vessels

mean arterial pressure (MAP)
: average driving force of blood to the tissues; approximated by taking diastolic pressure and adding 1/3 of pulse pressure

pulse
: alternating expansion and recoil of an artery as blood moves through the vessel; an indicator of heart rate

pulse pressure
: difference between the systolic and diastolic pressures

resistance
: any condition or parameter that slows or counteracts the flow of blood

respiratory pump
: increase in the volume of the thorax during inhalation that decreases air pressure, enabling venous blood to flow into the thoracic region, then exhalation increases pressure, moving blood into the atria

skeletal muscle pump
: effect on increasing blood pressure within veins by compression of the vessel caused by the contraction of nearby skeletal muscle

sphygmomanometer
: blood pressure cuff attached to a device that measures blood pressure

systolic pressure
: larger number recorded when measuring arterial blood pressure; represents the maximum value following ventricular contraction

vascular tone
: contractile state of smooth muscle in a blood vessel

Introduction to Fluid Dynamics and Its Biological and Medical Applications

class="introduction"

https://www.openstaxcollege.org/l/02planewing

In our daily lives, we often encounter fluids in motion: we turn on the faucet and water comes out; on a winter day, we watch smoke rise from a chimney. We use the power of water to produce electricity in hydroelectric dams. When constructing and managing them, engineers must consider the speed, volume, and power of the water flowing through the intakes and against the turbines.

In this chapter, you will learn about the behavior of fluids in motion, including rate of flow of a fluid, laminar and turbulent flow, the Bernoulli Effect and its applications, viscosity, and other topics in fluid dynamics. By the end of this chapter, you will learn how to calculate the rate of flow of a fluid, the velocity and pressure of air in Bernoulli Effect problems, the motion of an object in a viscous fluid, and Reynolds numbers that describe the onset of turbulence.

Flow Rate and Its Relation to Velocity

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Calculate flow rate.
- Define units of volume.
- Describe incompressible fluids.
- Explain the consequences of the equation of continuity.

The information presented in this section supports the following AP® learning objectives and science practices:

- **5.F.1.1** The student is able to make calculations of quantities related to flow of a fluid, using mass conservation principles (the continuity equation). **(S.P. 6.4, 7.2)**

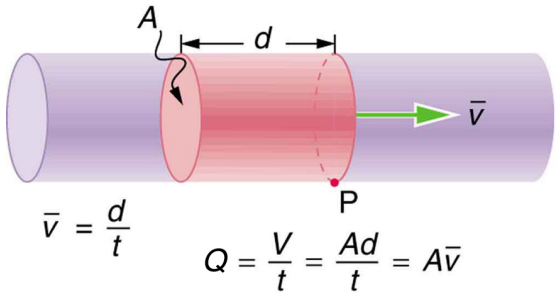
Flow rate Q is defined to be the volume of fluid passing by some location through an area during a period of time, as seen in [link]. In symbols, this can be written as

Equation:

$$Q = \frac{V}{t},$$

where V is the volume and t is the elapsed time.

The SI unit for flow rate is m^3/s, but a number of other units for Q are in common use. For example, the heart of a resting adult pumps blood at a rate of 5.00 liters per minute (L/min). Note that a **liter** (L) is 1/1000 of a cubic meter or 1000 cubic centimeters (10^{-3} m^3 or 10^3 cm^3). In this text we shall use whatever metric units are most convenient for a given situation.



Flow rate is the volume of fluid per unit time flowing past a point through the area A. Here the shaded cylinder of fluid flows past point P in a uniform pipe in time t. The volume of the cylinder is Ad and the average velocity is $v = d/t$ so that the flow rate is $Q = \mathrm{Ad}/t = Av$.

Example:
Calculating Volume from Flow Rate: The Heart Pumps a Lot of Blood in a Lifetime
How many cubic meters of blood does the heart pump in a 75-year lifetime, assuming the average flow rate is 5.00 L/min?
Strategy
Time and flow rate Q are given, and so the volume V can be calculated from the definition of flow rate.
Solution

Solving $Q = V/t$ for volume gives
Equation:

$$V = \mathrm{Qt}.$$

Substituting known values yields
Equation:

$$\begin{aligned} V &= \left(\tfrac{5.00\ \mathrm{L}}{1\ \mathrm{min}}\right)(75\ \mathrm{y})\left(\tfrac{1\ \mathrm{m}^3}{10^3\ \mathrm{L}}\right)\left(5.26 \times 10^5\ \tfrac{\mathrm{min}}{\mathrm{y}}\right) \\ &= 2.0 \times 10^5\ \mathrm{m}^3. \end{aligned}$$

Discussion
This amount is about 200,000 tons of blood. For comparison, this value is equivalent to about 200 times the volume of water contained in a 6-lane 50-m lap pool.

Flow rate and velocity are related, but quite different, physical quantities. To make the distinction clear, think about the flow rate of a river. The greater the velocity of the water, the greater the flow rate of the river. But flow rate also depends on the size of the river. A rapid mountain stream carries far less water than the Amazon River in Brazil, for example. The precise relationship between flow rate Q and velocity v is
Equation:

$$Q = Av,$$

where A is the cross-sectional area and v is the average velocity. This equation seems logical enough. The relationship tells us that flow rate is directly proportional to both the magnitude of the average velocity (hereafter referred to as the speed) and the size of a river, pipe, or other conduit. The larger the conduit, the greater its cross-sectional area. [link] illustrates how this relationship is obtained. The shaded cylinder has a volume
Equation:

$$V = \mathrm{Ad},$$

which flows past the point P in a time t. Dividing both sides of this relationship by t gives

Equation:

$$\frac{V}{t} = \frac{\mathrm{Ad}}{t}.$$

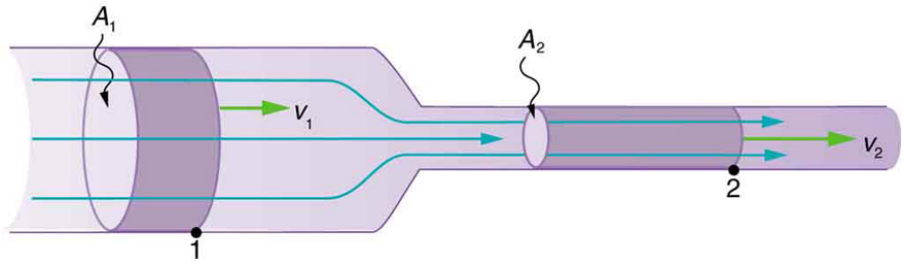
We note that $Q = V/t$ and the average speed is $v = d/t$. Thus the equation becomes $Q = Av$.

[link] shows an incompressible fluid flowing along a pipe of decreasing radius. Because the fluid is incompressible, the same amount of fluid must flow past any point in the tube in a given time to ensure continuity of flow. In this case, because the cross-sectional area of the pipe decreases, the velocity must necessarily increase. This logic can be extended to say that the flow rate must be the same at all points along the pipe. In particular, for points 1 and 2,

Equation:

$$\left.\begin{aligned} Q_1 &= Q_2 \\ A_1 v_1 &= A_2 v_2 \end{aligned}\right\}.$$

This is called the equation of continuity and is valid for any incompressible fluid. The consequences of the equation of continuity can be observed when water flows from a hose into a narrow spray nozzle: it emerges with a large speed—that is the purpose of the nozzle. Conversely, when a river empties into one end of a reservoir, the water slows considerably, perhaps picking up speed again when it leaves the other end of the reservoir. In other words, speed increases when cross-sectional area decreases, and speed decreases when cross-sectional area increases.



When a tube narrows, the same volume occupies a greater length. For the same volume to pass points 1 and 2 in a given time, the speed must be greater at point 2. The process is exactly reversible. If the fluid flows in the opposite direction, its speed will decrease when the tube widens. (Note that the relative volumes of the two cylinders and the corresponding velocity vector arrows are not drawn to scale.)

Since liquids are essentially incompressible, the equation of continuity is valid for all liquids. However, gases are compressible, and so the equation must be applied with caution to gases if they are subjected to compression or expansion.

Note:

Making Connections: Incompressible Fluid

The continuity equation tells us that the flow rate must be the same throughout an incompressible fluid. The flow rate, Q, has units of volume per unit time (m^3/s). Another way to think about it would be as a conservation principle, that the volume of fluid flowing past any point in a given amount of time must be conserved throughout the fluid.

For incompressible fluids, we can also say that the mass flowing past any point in a given amount of time must also be conserved. That is because the mass of a given volume of fluid is just the density of the fluid multiplied by the volume:

Equation:

$$m = \rho V$$

When we say a fluid is incompressible, we mean that the density of the fluid does not change. Every cubic meter of fluid has the same number of particles. There is no room to add more particles, nor is the fluid allowed to expand so that the particles will spread out. Since the density is constant, we can express the conservation principle as follows for any two regions of fluid flow, starting with the continuity equation:

Equation:

$$\frac{V_1}{t_1} = \frac{V_2}{t_2}$$

Equation:

$$\frac{\rho V_1}{t_1} = \frac{\rho V_2}{t_2}$$

Equation:

$$\frac{m_1}{t_1} = \frac{m_2}{t_2}$$

More generally, we say that the mass flow rate $\left(\frac{\Delta m}{\Delta t}\right)$ is conserved.

Example:

Calculating Fluid Speed: Speed Increases When a Tube Narrows

A nozzle with a radius of 0.250 cm is attached to a garden hose with a radius of 0.900 cm. The flow rate through hose and nozzle is 0.500 L/s. Calculate the speed of the water (a) in the hose and (b) in the nozzle.

Strategy

We can use the relationship between flow rate and speed to find both velocities. We will use the subscript 1 for the hose and 2 for the nozzle.

Solution for (a)

First, we solve $Q = Av$ for v_1 and note that the cross-sectional area is $A = \pi r^2$, yielding

Equation:

$$v_1 = \frac{Q}{A_1} = \frac{Q}{\pi r_1^2}.$$

Substituting known values and making appropriate unit conversions yields

Equation:

$$v_1 = \frac{(0.500\ \text{L/s})(10^{-3}\ \text{m}^3/\text{L})}{\pi(9.00 \times 10^{-3}\ \text{m})^2} = 1.96\ \text{m/s}.$$

Solution for (b)

We could repeat this calculation to find the speed in the nozzle v_2, but we will use the equation of continuity to give a somewhat different insight. Using the equation which states

Equation:

$$A_1 v_1 = A_2 v_2,$$

solving for v_2 and substituting πr^2 for the cross-sectional area yields

Equation:

$$v_2 = \frac{A_1}{A_2} v_1 = \frac{\pi r_1^2}{\pi r_2^2} v_1 = \frac{r_{1^2}}{r_{2^2}} v_1.$$

Substituting known values,

Equation:

$$v_2 = \frac{(0.900\ \text{cm})^2}{(0.250\ \text{cm})^2} 1.96\ \text{m/s} = 25.5\ \text{m/s}.$$

Discussion

A speed of 1.96 m/s is about right for water emerging from a nozzleless hose. The nozzle produces a considerably faster stream merely by constricting the flow to a narrower tube.

Note:

Making Connections: Different-Sized Pipes

For incompressible fluids, the density of the fluid remains constant throughout, no matter the flow rate or the size of the opening through which the fluid flows. We say that, to ensure continuity of flow, the amount of fluid that flows past any point is constant. That amount can be measured by either volume or mass.

Flow rate has units of volume/time (m^3/s or L/s). Mass flow rate $\left(\frac{\Delta m}{\Delta t}\right)$ has units of mass/time (kg/s) and can be calculated from the flow rate by using the density:

Equation:

$$m = \rho V$$

The average mass flow rate can be found from the flow rate:

Equation:

$$\frac{\Delta m}{\Delta t} = \frac{m}{t} = \frac{\dot{\rho V}}{t} = \rho Q = \rho A v$$

Suppose that crude oil with a density of 880 kg/m^3 is flowing through a pipe with a diameter of 55 cm and a speed of 1.8 m/s. Calculate the new speed of the crude oil when the pipe narrows to a new diameter of 31 cm, and calculate the mass flow rate in both sections of the pipe, assuming the density of the oil is constant throughout the pipe.

Solution: To calculate the new speed, we simply use the continuity equation.

Since the cross section of a pipe is a circle, the area of each cross section can be found as follows:

For the larger pipe:

Equation:

$$A_1 = \pi\left(\frac{d_1}{2}\right)^2 = \pi(0.275)^2 = 0.238\,\mathrm{m}^2$$

For the smaller pipe:

Equation:

$$A_2 = \pi(0.155)^2 = \ 0.0755\,\mathrm{m}^2$$

So the larger part of the pipe (A_1) has a cross-sectional area of 0.238 m^2, and the smaller part of the pipe (A_2) has a cross-sectional area of 0.0755 m^2. The continuity equation tells us that the oil will flow faster through the portion of the pipe with the smaller cross-sectional area. Using the continuity equation, we get

Equation:

$$A_1 v_1 = \ A_2 v_2$$

Equation:

$$v_2 = \ \left(\frac{A_1}{A_2}\right) v_1 = \ \left(\frac{0.238}{0.0755}\right)(1.8) = \ 5.7\,\mathrm{m/s}$$

So we find that the oil is flowing at a speed of 1.8 m/s through the larger section of the pipe (A_1), and it is flowing much faster (5.7 m/s) through the smaller section (A_2).

The mass flow rate in both sections should be the same.

For the larger portion of the pipe:

Equation:

$$\left(\frac{\Delta m}{\Delta t}\right)_1 = \ \rho A_1 v_1 = (880)\,(0.238)\,(1.8) = \ 380\,\mathrm{kg/s}$$

For the smaller portion of the pipe:

Equation:

$$\left(\frac{\Delta m}{\Delta t}\right)_2 = \ \rho A_2 v_2 = (880)\,(0.75538)\,(5.7) = \ 380\,\mathrm{kg/s}$$

And so mass is conserved throughout the pipe. Every second, 380 kg of oil flows out of the larger portion of the pipe, and 380 kg of oil flows into the smaller portion of the pipe.

The solution to the last part of the example shows that speed is inversely proportional to the *square* of the radius of the tube, making for large effects when radius varies. We can blow out a candle at quite a distance, for example, by pursing our lips, whereas blowing on a candle with our mouth wide open is quite ineffective.

In many situations, including in the cardiovascular system, branching of the flow occurs. The blood is pumped from the heart into arteries that subdivide into smaller arteries (arterioles) which branch into very fine vessels called capillaries. In this situation, continuity of flow is maintained but it is the *sum* of the flow rates in each of the branches in any portion along the tube that is maintained. The equation of continuity in a more general form becomes

Equation:

$$n_1 A_1 v_1 = n_2 A_2 v_2,$$

where n_1 and n_2 are the number of branches in each of the sections along the tube.

Example:

Calculating Flow Speed and Vessel Diameter: Branching in the Cardiovascular System

The aorta is the principal blood vessel through which blood leaves the heart in order to circulate around the body. (a) Calculate the average speed of the blood in the aorta if the flow rate is 5.0 L/min. The aorta has a radius of 10 mm. (b) Blood also flows through smaller blood vessels known as capillaries. When the rate of blood flow in the aorta is 5.0 L/min, the speed of blood in the capillaries is about 0.33 mm/s. Given that the average diameter of a capillary is 8.0 μm, calculate the number of capillaries in the blood circulatory system.

Strategy

We can use $Q = Av$ to calculate the speed of flow in the aorta and then use the general form of the equation of continuity to calculate the number of capillaries as all of the other variables are known.

Solution for (a)
The flow rate is given by $Q = Av$ or $v = \frac{Q}{\pi r^2}$ for a cylindrical vessel. Substituting the known values (converted to units of meters and seconds) gives
Equation:

$$v = \frac{(5.0\ \text{L/min})(10^{-3}\ \text{m}^3/\text{L})(1\ \text{min}/60\ \text{s})}{\pi(0.010\ \text{m})^2} = 0.27\ \text{m/s}.$$

Solution for (b)
Using $n_1 A_1 v_1 = n_2 A_2 v_1$, assigning the subscript 1 to the aorta and 2 to the capillaries, and solving for n_2 (the number of capillaries) gives $n_2 = \frac{n_1 A_1 v_1}{A_2 v_2}$. Converting all quantities to units of meters and seconds and substituting into the equation above gives
Equation:

$$n_2 = \frac{(1)(\pi)\left(10 \times 10^{-3}\ \text{m}\right)^2 (0.27\ \text{m/s})}{(\pi)\left(4.0 \times 10^{-6}\ \text{m}\right)^2 \left(0.33 \times 10^{-3}\ \text{m/s}\right)} = 5.0 \times 10^9\ \text{capillaries}.$$

Discussion
Note that the speed of flow in the capillaries is considerably reduced relative to the speed in the aorta due to the significant increase in the total cross-sectional area at the capillaries. This low speed is to allow sufficient time for effective exchange to occur although it is equally important for the flow not to become stationary in order to avoid the possibility of clotting. Does this large number of capillaries in the body seem reasonable? In active muscle, one finds about 200 capillaries per mm^3, or about 200×10^6 per 1 kg of muscle. For 20 kg of muscle, this amounts to about 4×10^9 capillaries.

Note:
Making Connections: Syringes

A horizontally oriented hypodermic syringe has a barrel diameter of 1.2 cm and a needle diameter of 2.4 mm. A plunger pushes liquid in the barrel at a rate of 4.0 mm/s. Calculate the flow rate of liquid in both parts of the syringe (in mL/s) and the velocity of the liquid emerging from the needle.
Solution:
First, calculate the area of both parts of the syringe:
Equation:

$$A_1 = \pi\left(\frac{d_1}{2}\right)^2 = \pi(0.006)^2 = 1.13 \times 10^{-4}\,\mathrm{m}^2$$

Equation:

$$A_2 = \pi\left(\frac{d_2}{2}\right)^2 = \pi(0.0012)^2 = 4.52 \times 10^{-6}\,\mathrm{m}^2$$

Next, we can use the continuity equation to find the velocity of the liquid in the smaller part of the barrel (v_2):
Equation:

$$A_1 v_1 = A_2 v_2$$

Equation:

$$v_2 = \left(\frac{A_1}{A_2}\right) v_1$$

Equation:

$$v_2 = \left(\frac{1.13\times10^{-4}}{4.52\times10^{-6}}\right)(0.004) = 0.10\,\mathrm{m/s}$$

Double-check the numbers to be sure that the flow rate in both parts of the syringe is the same:
Equation:

$$Q_1 = A_1 v_1 = \left(1.13\times10^{-4}\right)(0.004) = 4.52\times10^{-7}\,\mathrm{m^3/s}$$

Equation:

$$Q_2 = \ A_2 v_2 = \left(4.52 \times 10^{-6}\right)(0.10) = \ 4.52 \times 10^{-7}\ \mathrm{m^3/s}$$

Finally, by converting to mL/s:
Equation:

$$\left(\frac{4.52 \times 10^{-7}\ \mathrm{m^3}}{1\ \mathrm{s}}\right)\left(\frac{10^6\ \mathrm{mL}}{1\ m^3}\right) = 0.452\ \mathrm{mL/s}$$

Section Summary

- Flow rate Q is defined to be the volume V flowing past a point in time t, or $Q = \frac{V}{t}$ where V is volume and t is time.
- The SI unit of volume is $\mathrm{m^3}$.
- Another common unit is the liter (L), which is $10^{-3}\ \mathrm{m^3}$.
- Flow rate and velocity are related by $Q = Av$ where A is the cross-sectional area of the flow and v is its average velocity.
- For incompressible fluids, flow rate at various points is constant. That is,
  Equation:

$$\begin{aligned} Q_1 &= Q_2 \\ A_1 v_1 &= A_2 v_2 \\ n_1 A_1 v_1 &= n_2 A_2 v_2 \end{aligned} \quad .$$

Conceptual Questions

Exercise:

Problem:

What is the difference between flow rate and fluid velocity? How are they related?

Exercise:

Problem:

Many figures in the text show streamlines. Explain why fluid velocity is greatest where streamlines are closest together. (Hint: Consider the relationship between fluid velocity and the cross-sectional area through which it flows.)

Exercise:

Problem:

Identify some substances that are incompressible and some that are not.

Problems & Exercises

Exercise:

Problem:

What is the average flow rate in cm^3/s of gasoline to the engine of a car traveling at 100 km/h if it averages 10.0 km/L?

Solution:

$2.78\ \text{cm}^3/\text{s}$

Exercise:

Problem:

The heart of a resting adult pumps blood at a rate of 5.00 L/min. (a) Convert this to cm^3/s. (b) What is this rate in m^3/s?

Exercise:

Problem:

Blood is pumped from the heart at a rate of 5.0 L/min into the aorta (of radius 1.0 cm). Determine the speed of blood through the aorta.

Solution:

27 cm/s

Exercise:

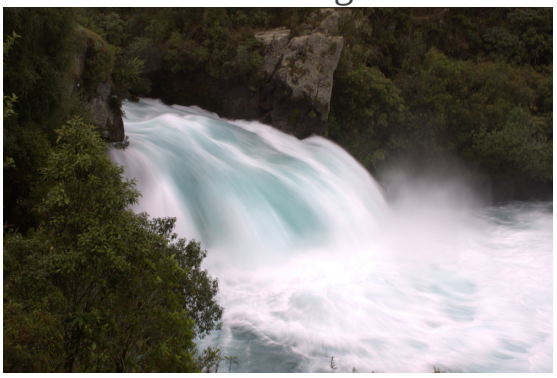
Problem:

Blood is flowing through an artery of radius 2 mm at a rate of 40 cm/s. Determine the flow rate and the volume that passes through the artery in a period of 30 s.

Exercise:

Problem:

The Huka Falls on the Waikato River is one of New Zealand's most visited natural tourist attractions (see [link]). On average the river has a flow rate of about 300,000 L/s. At the gorge, the river narrows to 20 m wide and averages 20 m deep. (a) What is the average speed of the river in the gorge? (b) What is the average speed of the water in the river downstream of the falls when it widens to 60 m and its depth increases to an average of 40 m?



The Huka Falls in Taupo,

New Zealand, demonstrate flow rate. (credit: RaviGogna, Flickr)

Solution:

(a) 0.75 m/s

(b) 0.13 m/s

Exercise:

Problem:

A major artery with a cross-sectional area of $1.00\ \text{cm}^2$ branches into 18 smaller arteries, each with an average cross-sectional area of $0.400\ \text{cm}^2$. By what factor is the average velocity of the blood reduced when it passes into these branches?

Exercise:

Problem:

(a) As blood passes through the capillary bed in an organ, the capillaries join to form venules (small veins). If the blood speed increases by a factor of 4.00 and the total cross-sectional area of the venules is $10.0\ \text{cm}^2$, what is the total cross-sectional area of the capillaries feeding these venules? (b) How many capillaries are involved if their average diameter is $10.0\ \mu\text{m}$?

Solution:

(a) $40.0\ \text{cm}^2$

(b) 5.09×10^7

Exercise:

Problem:

The human circulation system has approximately 1×10^9 capillary vessels. Each vessel has a diameter of about $8\ \mu\text{m}$. Assuming cardiac output is 5 L/min, determine the average velocity of blood flow through each capillary vessel.

Exercise:

Problem:

(a) Estimate the time it would take to fill a private swimming pool with a capacity of 80,000 L using a garden hose delivering 60 L/min. (b) How long would it take to fill if you could divert a moderate size river, flowing at $5000\ \text{m}^3/\text{s}$, into it?

Solution:

(a) 22 h

(b) 0.016 s

Exercise:

Problem:

The flow rate of blood through a 2.00×10^{-6}-m -radius capillary is $3.80 \times 10^9\ \text{cm}^3/\text{s}^2$. (a) What is the speed of the blood flow? (This small speed allows time for diffusion of materials to and from the blood.) (b) Assuming all the blood in the body passes through capillaries, how many of them must there be to carry a total flow of $90.0\ \text{cm}^3/\text{s}$? (The large number obtained is an overestimate, but it is still reasonable.)

Exercise:

Problem:

(a) What is the fluid speed in a fire hose with a 9.00-cm diameter carrying 80.0 L of water per second? (b) What is the flow rate in cubic meters per second? (c) Would your answers be different if salt water replaced the fresh water in the fire hose?

Solution:

(a) 12.6 m/s

(b) $0.0800\ \mathrm{m}^3/\mathrm{s}$

(c) No, independent of density.

Exercise:

Problem:

The main uptake air duct of a forced air gas heater is 0.300 m in diameter. What is the average speed of air in the duct if it carries a volume equal to that of the house's interior every 15 min? The inside volume of the house is equivalent to a rectangular solid 13.0 m wide by 20.0 m long by 2.75 m high.

Exercise:

Problem:

Water is moving at a velocity of 2.00 m/s through a hose with an internal diameter of 1.60 cm. (a) What is the flow rate in liters per second? (b) The fluid velocity in this hose's nozzle is 15.0 m/s. What is the nozzle's inside diameter?

Solution:

(a) 0.402 L/s

(b) 0.584 cm

Exercise:

Problem:

Prove that the speed of an incompressible fluid through a constriction, such as in a Venturi tube, increases by a factor equal to the square of the factor by which the diameter decreases. (The converse applies for flow out of a constriction into a larger-diameter region.)

Exercise:

Problem:

Water emerges straight down from a faucet with a 1.80-cm diameter at a speed of 0.500 m/s. (Because of the construction of the faucet, there is no variation in speed across the stream.) (a) What is the flow rate in cm^3/s? (b) What is the diameter of the stream 0.200 m below the faucet? Neglect any effects due to surface tension.

Solution:

(a) $127\ \text{cm}^3/\text{s}$

(b) 0.890 cm

Exercise:

Problem: Unreasonable Results

A mountain stream is 10.0 m wide and averages 2.00 m in depth. During the spring runoff, the flow in the stream reaches $100{,}000\ \text{m}^3/\text{s}$. (a) What is the average velocity of the stream under these conditions? (b) What is unreasonable about this velocity? (c) What is unreasonable or inconsistent about the premises?

Test Prep for AP Courses

Exercise:

Problem:

Water flows through a small horizontal pipe with a speed of 12 m/s into a larger part of the pipe for which the diameter of the pipe is doubled. What is the speed of the water in the larger part of the pipe?

a. 0.75 m/s
b. 3.0 m/s
c. 6.0 m/s
d. 12 m/s

Solution:

(c)

Exercise:

Problem:

A popular pool toy allows the user to push a plunger to compress water in a barrel with a diameter of 3.0 cm. The water is compressed with a speed of 1.2 m/s and emerges from a small opening with a speed of 15 m/s. What is the diameter of the opening? Assume the toy is oriented horizontally.

Glossary

flow rate
: abbreviated Q, it is the volume V that flows past a particular point during a time t, or $Q = V/t$

liter
: a unit of volume, equal to 10^{-3} m^3

Viscosity and Laminar Flow; Poiseuille’s Law

- Define laminar flow and turbulent flow.
- Explain what viscosity is.
- Calculate flow and resistance with Poiseuille’s law.
- Explain how pressure drops due to resistance.

Laminar Flow and Viscosity

When you pour yourself a glass of juice, the liquid flows freely and quickly. But when you pour syrup on your pancakes, that liquid flows slowly and sticks to the pitcher. The difference is fluid friction, both within the fluid itself and between the fluid and its surroundings. We call this property of fluids *viscosity*. Juice has low viscosity, whereas syrup has high viscosity. In the previous sections we have considered ideal fluids with little or no viscosity. In this section, we will investigate what factors, including viscosity, affect the rate of fluid flow.

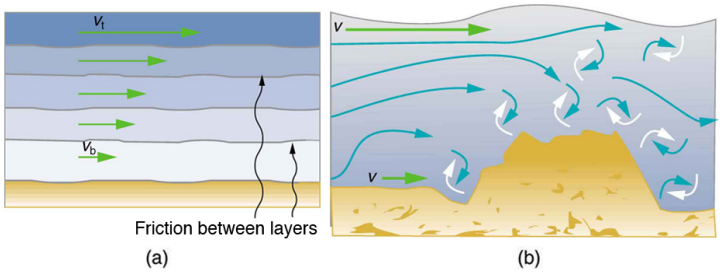
The precise definition of viscosity is based on *laminar*, or nonturbulent, flow. Before we can define viscosity, then, we need to define laminar flow and turbulent flow. [link] shows both types of flow. **Laminar** flow is characterized by the smooth flow of the fluid in layers that do not mix. Turbulent flow, or **turbulence**, is characterized by eddies and swirls that mix layers of fluid together.



Smoke rises smoothly for a while and then

begins to form swirls and eddies. The smooth flow is called laminar flow, whereas the swirls and eddies typify turbulent flow. If you watch the smoke (being careful not to breathe on it), you will notice that it rises more rapidly when flowing smoothly than after it becomes turbulent, implying that turbulence poses more resistance to flow. (credit: Creativity103)

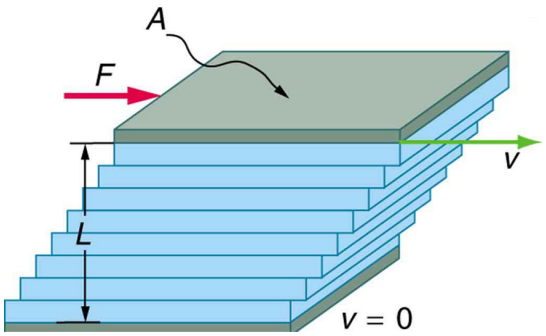
[link] shows schematically how laminar and turbulent flow differ. Layers flow without mixing when flow is laminar. When there is turbulence, the layers mix, and there are significant velocities in directions other than the overall direction of flow. The lines that are shown in many illustrations are the paths followed by small volumes of fluids. These are called *streamlines*. Streamlines are smooth and continuous when flow is laminar, but break up and mix when flow is turbulent. Turbulence has two main causes. First, any obstruction or sharp corner, such as in a faucet, creates turbulence by imparting velocities perpendicular to the flow. Second, high speeds cause turbulence. The drag both between adjacent layers of fluid and between the fluid and its surroundings forms swirls and eddies, if the speed is great enough. We shall concentrate on laminar flow for the remainder of this section, leaving certain aspects of turbulence for later sections.



(a) Laminar flow occurs in layers without mixing. Notice that viscosity causes drag between layers as well as with the fixed surface. (b) An obstruction in the vessel produces turbulence. Turbulent flow mixes the fluid. There is more interaction, greater heating, and more resistance than in laminar flow.

Note:
Making Connections: Take-Home Experiment: Go Down to the River
Try dropping simultaneously two sticks into a flowing river, one near the edge of the river and one near the middle. Which one travels faster? Why?

[link] shows how viscosity is measured for a fluid. Two parallel plates have the specific fluid between them. The bottom plate is held fixed, while the top plate is moved to the right, dragging fluid with it. The layer (or lamina) of fluid in contact with either plate does not move relative to the plate, and so the top layer moves at v while the bottom layer remains at rest. Each successive layer from the top down exerts a force on the one below it, trying to drag it along, producing a continuous variation in speed from v to 0 as shown. Care is taken to insure that the flow is laminar; that is, the layers do not mix. The motion in [link] is like a continuous shearing motion. Fluids have zero shear strength, but the *rate* at which they are sheared is related to the same geometrical factors A and L as is shear deformation for solids.



The graphic shows laminar flow of fluid between two plates of area A. The bottom plate is fixed. When the top plate is pushed to the right, it drags the fluid along with it.

A force F is required to keep the top plate in [link] moving at a constant velocity v, and experiments have shown that this force depends on four factors. First, F is directly proportional to v (until the speed is so high that turbulence occurs—then a much larger force is needed, and it has a more complicated dependence on v). Second, F is proportional to the area A of the plate. This relationship seems reasonable, since A is directly proportional to the amount of fluid being moved. Third, F is inversely proportional to the distance between the plates L. This relationship is also reasonable; L is like a lever arm, and the greater the lever arm, the less force that is needed. Fourth, F is directly proportional to *the coefficient of viscosity*, η. The greater the viscosity, the greater the force required. These dependencies are combined into the equation

Equation:

$$F = \eta \frac{\mathrm{vA}}{L},$$

which gives us a working definition of fluid **viscosity** η. Solving for η gives

Equation:

$$\eta = \frac{\mathrm{FL}}{\mathrm{vA}},$$

which defines viscosity in terms of how it is measured. The SI unit of viscosity is $\mathrm{N \cdot m/[(m/s)m^2] = (N/m^2)s}$ or $\mathrm{Pa \cdot s}$. [link] lists the coefficients of viscosity for various fluids.

Viscosity varies from one fluid to another by several orders of magnitude. As you might expect, the viscosities of gases are much less than those of liquids, and these viscosities are often temperature dependent. The viscosity of blood can be reduced by aspirin consumption, allowing it to flow more easily around the body. (When used over the long term in low doses, aspirin can help prevent heart attacks, and reduce the risk of blood clotting.)

Laminar Flow Confined to Tubes—Poiseuille's Law

What causes flow? The answer, not surprisingly, is pressure difference. In fact, there is a very simple relationship between horizontal flow and pressure. Flow rate Q is in the direction from high to low pressure. The greater the pressure differential between two points, the greater the flow rate. This relationship can be stated as
Equation:

$$Q = \frac{P_2 - P_1}{R},$$

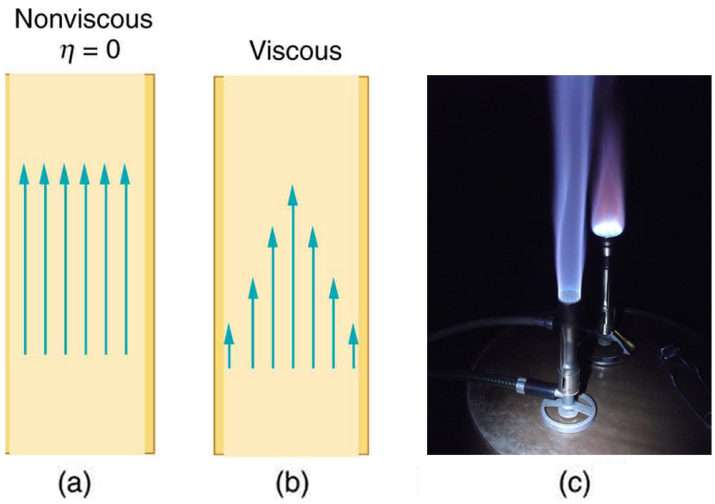
where P_1 and P_2 are the pressures at two points, such as at either end of a tube, and R is the resistance to flow. The resistance R includes everything, except pressure, that affects flow rate. For example, R is greater for a long tube than for a short one. The greater the viscosity of a fluid, the greater the value of R. Turbulence greatly increases R, whereas increasing the diameter of a tube decreases R.

If viscosity is zero, the fluid is frictionless and the resistance to flow is also zero. Comparing frictionless flow in a tube to viscous flow, as in [link], we see that for a viscous fluid, speed is greatest at midstream because of drag at the boundaries. We can see the effect of viscosity in a Bunsen burner flame, even though the viscosity of natural gas is small.

The resistance R to laminar flow of an incompressible fluid having viscosity η through a horizontal tube of uniform radius r and length l, such as the one in [link], is given by
Equation:

$$R = \frac{8\eta l}{\pi r^4}.$$

This equation is called **Poiseuille's law for resistance** after the French scientist J. L. Poiseuille (1799–1869), who derived it in an attempt to understand the flow of blood, an often turbulent fluid.



(a) If fluid flow in a tube has negligible resistance, the speed is the same all across the tube. (b) When a viscous fluid flows through a tube, its speed at the walls is zero, increasing steadily to its maximum at the center of the tube. (c) The shape of the Bunsen burner flame is due to the velocity profile across the tube. (credit: Jason Woodhead)

Let us examine Poiseuille's expression for R to see if it makes good intuitive sense. We see that resistance is directly proportional to both fluid viscosity η and the length l of a tube. After all, both of these directly affect the amount of friction encountered —the greater either is, the greater the resistance and the smaller the flow. The radius r of a tube affects the resistance, which again makes sense, because the greater the radius, the greater the flow (all other factors remaining the same). But it is surprising that r is raised to the *fourth* power in Poiseuille's law. This exponent means that any change in the radius of a tube has a very large effect on resistance. For example, doubling the radius of a tube decreases resistance by a factor of $2^4 = 16$.

Taken together, $Q = \frac{P_2 - P_1}{R}$ and $R = \frac{8\eta l}{\pi r^4}$ give the following expression for flow rate:

Equation:

$$Q = \frac{(P_2 - P_1)\pi r^4}{8\eta l}.$$

This equation describes laminar flow through a tube. It is sometimes called Poiseuille's law for laminar flow, or simply **Poiseuille's law**.

Example:

Using Flow Rate: Plaque Deposits Reduce Blood Flow

Suppose the flow rate of blood in a coronary artery has been reduced to half its normal value by plaque deposits. By what factor has the radius of the artery been reduced, assuming no turbulence occurs?

Strategy

Assuming laminar flow, Poiseuille's law states that

Equation:

$$Q = \frac{(P_2 - P_1)\pi r^4}{8\eta l}.$$

We need to compare the artery radius before and after the flow rate reduction.

Solution

With a constant pressure difference assumed and the same length and viscosity, along the artery we have

Equation:

$$\frac{Q_1}{r_1^4} = \frac{Q_2}{r_2^4}.$$

So, given that $Q_2 = 0.5Q_1$, we find that $r_2^4 = 0.5r_1^4$.

Therefore, $r_2 = (0.5)^{0.25} r_1 = 0.841 r_1$, a decrease in the artery radius of 16%.

Discussion

This decrease in radius is surprisingly small for this situation. To restore the blood flow in spite of this buildup would require an increase in the pressure difference $(P_2 - P_1)$ of a factor of two, with subsequent strain on the heart.

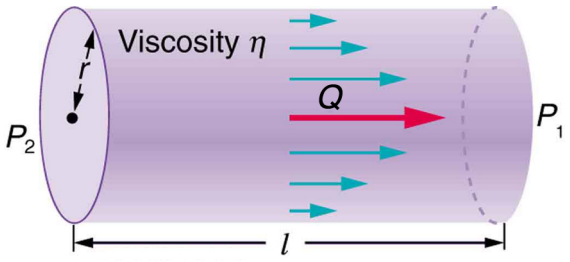
Fluid	Temperature (°C)	Viscosity η (mPa·s)
Gases		
Air	0	0.0171
	20	0.0181
	40	0.0190
	100	0.0218
Ammonia	20	0.00974
Carbon dioxide	20	0.0147
Helium	20	0.0196
Hydrogen	0	0.0090
Mercury	20	0.0450
Oxygen	20	0.0203
Steam	100	0.0130
Liquids		
Water	0	1.792
	20	1.002
	37	0.6947
	40	0.653
	100	0.282
Whole blood[footnote]	20	3.015

Fluid	Temperature (ºC)	Viscosity η (mPa·s)
The ratios of the viscosities of blood to water are nearly constant between 0°C and 37°C.	37	2.084
Blood plasma[footnote] See note on Whole Blood.	20	1.810
	37	1.257
Ethyl alcohol	20	1.20
Methanol	20	0.584
Oil (heavy machine)	20	660
Oil (motor, SAE 10)	30	200
Oil (olive)	20	138
Glycerin	20	1500
Honey	20	2000–10000
Maple Syrup	20	2000–3000
Milk	20	3.0
Oil (Corn)	20	65

Coefficients of Viscosity of Various Fluids

The circulatory system provides many examples of Poiseuille's law in action—with blood flow regulated by changes in vessel size and blood pressure. Blood vessels are not rigid but elastic. Adjustments to blood flow are primarily made by varying the size of the vessels, since the resistance is so sensitive to the radius. During vigorous exercise, blood vessels are selectively dilated to important muscles and organs and blood pressure increases. This creates both greater overall blood flow and increased flow to specific areas. Conversely, decreases in vessel radii, perhaps from plaques in

the arteries, can greatly reduce blood flow. If a vessel's radius is reduced by only 5% (to 0.95 of its original value), the flow rate is reduced to about $(0.95)^4 = 0.81$ of its original value. A 19% decrease in flow is caused by a 5% decrease in radius. The body may compensate by increasing blood pressure by 19%, but this presents hazards to the heart and any vessel that has weakened walls. Another example comes from automobile engine oil. If you have a car with an oil pressure gauge, you may notice that oil pressure is high when the engine is cold. Motor oil has greater viscosity when cold than when warm, and so pressure must be greater to pump the same amount of cold oil.



Poiseuille's law applies to laminar flow of an incompressible fluid of viscosity η through a tube of length l and radius r. The direction of flow is from greater to lower pressure. Flow rate Q is directly proportional to the pressure difference $P_2 - P_1$, and inversely proportional to the length l of the tube and viscosity η of the fluid. Flow rate increases with r^4, the fourth power of the radius.

Example:
What Pressure Produces This Flow Rate?

An intravenous (IV) system is supplying saline solution to a patient at the rate of $0.120\ \text{cm}^3/\text{s}$ through a needle of radius 0.150 mm and length 2.50 cm. What pressure is needed at the entrance of the needle to cause this flow, assuming the viscosity of the saline solution to be the same as that of water? The gauge pressure of the blood in the patient's vein is 8.00 mm Hg. (Assume that the temperature is 20^{o}C .)

Strategy

Assuming laminar flow, Poiseuille's law applies. This is given by

Equation:

$$Q = \frac{(P_2 - P_1)\pi r^4}{8\eta l},$$

where P_2 is the pressure at the entrance of the needle and P_1 is the pressure in the vein. The only unknown is P_2.

Solution

Solving for P_2 yields

Equation:

$$P_2 = \frac{8\eta l}{\pi r^4} Q + P_{1.}$$

P_1 is given as 8.00 mm Hg, which converts to $1.066 \times 10^3\ \text{N/m}^2$. Substituting this and the other known values yields

Equation:

$$\begin{aligned} P_2 &= \left[\frac{8(1.00\times10^{-3}\ \text{N}\cdot\text{s/m}^2)(2.50\times10^{-2}\ \text{m})}{\pi(0.150\times10^{-3}\ \text{m})^4}\right](1.20 \times 10^{-7}\ \text{m}^3/\text{s}) + 1.066 \times 10^3\ \text{N/m}^2 \\ &= 1.62 \times 10^4\ \text{N/m}^2. \end{aligned}$$

Discussion

This pressure could be supplied by an IV bottle with the surface of the saline solution 1.61 m above the entrance to the needle (this is left for you to solve in this chapter's Problems and Exercises), assuming that there is negligible pressure drop in the tubing leading to the needle.

Flow and Resistance as Causes of Pressure Drops

You may have noticed that water pressure in your home might be lower than normal on hot summer days when there is more use. This pressure drop occurs in the water

main before it reaches your home. Let us consider flow through the water main as illustrated in [link]. We can understand why the pressure P_1 to the home drops during times of heavy use by rearranging

Equation:

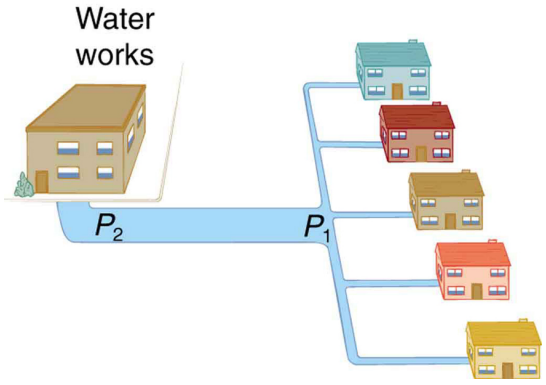
$$Q = \frac{P_2 - P_1}{R}$$

to

Equation:

$$P_2 - P_1 = RQ,$$

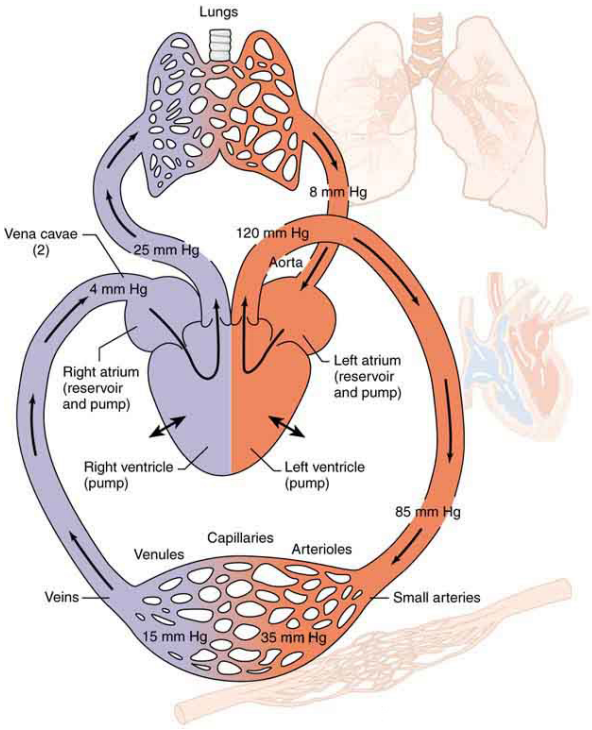
where, in this case, P_2 is the pressure at the water works and R is the resistance of the water main. During times of heavy use, the flow rate Q is large. This means that $P_2 - P_1$ must also be large. Thus P_1 must decrease. It is correct to think of flow and resistance as causing the pressure to drop from P_2 to P_1. $P_2 - P_1 = RQ$ is valid for both laminar and turbulent flows.



During times of heavy use, there is a significant pressure drop in a water main, and P_1 supplied to users is significantly less than P_2 created at the water works. If the flow is very small, then the pressure drop is negligible, and $P_2 \approx P_1$.

We can use $P_2 - P_1 = RQ$ to analyze pressure drops occurring in more complex systems in which the tube radius is not the same everywhere. Resistance will be much greater in narrow places, such as an obstructed coronary artery. For a given flow rate Q, the pressure drop will be greatest where the tube is most narrow. This is how water faucets control flow. Additionally, R is greatly increased by turbulence, and a constriction that creates turbulence greatly reduces the pressure downstream. Plaque in an artery reduces pressure and hence flow, both by its resistance and by the turbulence it creates.

[link] is a schematic of the human circulatory system, showing average blood pressures in its major parts for an adult at rest. Pressure created by the heart's two pumps, the right and left ventricles, is reduced by the resistance of the blood vessels as the blood flows through them. The left ventricle increases arterial blood pressure that drives the flow of blood through all parts of the body except the lungs. The right ventricle receives the lower pressure blood from two major veins and pumps it through the lungs for gas exchange with atmospheric gases – the disposal of carbon dioxide from the blood and the replenishment of oxygen. Only one major organ is shown schematically, with typical branching of arteries to ever smaller vessels, the smallest of which are the capillaries, and rejoining of small veins into larger ones. Similar branching takes place in a variety of organs in the body, and the circulatory system has considerable flexibility in flow regulation to these organs by the dilation and constriction of the arteries leading to them and the capillaries within them. The sensitivity of flow to tube radius makes this flexibility possible over a large range of flow rates.



Schematic of the circulatory system. Pressure difference is created by the two pumps in the heart and is reduced by resistance in the vessels. Branching of vessels into capillaries allows blood to reach individual cells and exchange substances, such as oxygen and waste products, with them. The system has an impressive ability to regulate flow to individual organs, accomplished largely by varying vessel diameters.

Each branching of larger vessels into smaller vessels increases the total cross-sectional area of the tubes through which the blood flows. For example, an artery with a cross section of 1 cm^2 may branch into 20 smaller arteries, each with cross sections of 0.5 cm^2, with a total of 10 cm^2. In that manner, the resistance of the branchings is reduced so that pressure is not entirely lost. Moreover, because $Q = Av$ and A increases through branching, the average velocity of the blood in the smaller vessels is reduced. The blood velocity in the aorta ($\mathrm{diameter} = 1$ cm) is about 25 cm/s, while in the capillaries (20μm in diameter) the velocity is about 1

mm/s. This reduced velocity allows the blood to exchange substances with the cells in the capillaries and alveoli in particular.

Section Summary

- Laminar flow is characterized by smooth flow of the fluid in layers that do not mix.
- Turbulence is characterized by eddies and swirls that mix layers of fluid together.
- Fluid viscosity η is due to friction within a fluid. Representative values are given in [link]. Viscosity has units of $(\mathrm{N/m^2})\mathrm{s}$ or $\mathrm{Pa \cdot s}$.
- Flow is proportional to pressure difference and inversely proportional to resistance:
  Equation:

$$Q = \frac{P_2 - P_1}{R}.$$

- For laminar flow in a tube, Poiseuille's law for resistance states that
  Equation:

$$R = \frac{8\eta l}{\pi r^4}.$$

- Poiseuille's law for flow in a tube is
  Equation:

$$Q = \frac{(P_2 - P_1)\pi r^4}{8\eta l}.$$

- The pressure drop caused by flow and resistance is given by
  Equation:

$$P_2 - P_1 = RQ.$$

Conceptual Questions

Exercise:

Problem:

Explain why the viscosity of a liquid decreases with temperature—that is, how might increased temperature reduce the effects of cohesive forces in a liquid? Also explain why the viscosity of a gas increases with temperature—that is, how does increased gas temperature create more collisions between atoms and molecules?

Exercise:

Problem:

When paddling a canoe upstream, it is wisest to travel as near to the shore as possible. When canoeing downstream, it may be best to stay near the middle. Explain why.

Exercise:

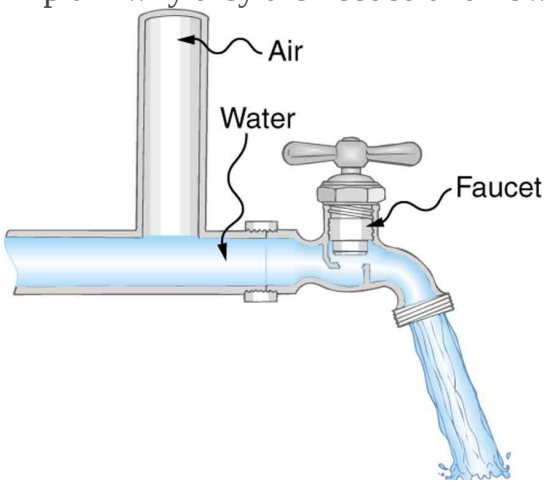
Problem:

Why does flow decrease in your shower when someone flushes the toilet?

Exercise:

Problem:

Plumbing usually includes air-filled tubes near water faucets, as shown in [link]. Explain why they are needed and how they work.



The vertical tube near the water tap remains full of air and serves a useful purpose.

Problems & Exercises

Exercise:

Problem:

(a) Calculate the retarding force due to the viscosity of the air layer between a cart and a level air track given the following information—air temperature is $20^{\circ}\ \mathrm{C}$, the cart is moving at 0.400 m/s, its surface area is $2.50 \times 10^{-2}\ \mathrm{m}^2$, and the thickness of the air layer is $6.00 \times 10^{-5}\ \mathrm{m}$. (b) What is the ratio of this force to the weight of the 0.300-kg cart?

Solution:

(a) $3.02 \times 10^{-3}\ \mathrm{N}$

(b) 1.03×10^{-3}

Exercise:

Problem:

What force is needed to pull one microscope slide over another at a speed of 1.00 cm/s, if there is a 0.500-mm-thick layer of $20^{\circ}\ \mathrm{C}$ water between them and the contact area is $8.00\ \mathrm{cm}^2$?

Exercise:

Problem:

A glucose solution being administered with an IV has a flow rate of $4.00\ \mathrm{cm}^3/\mathrm{min}$. What will the new flow rate be if the glucose is replaced by whole blood having the same density but a viscosity 2.50 times that of the glucose? All other factors remain constant.

Solution:

$1.60\ \mathrm{cm}^3/\mathrm{min}$

Exercise:

Problem:

The pressure drop along a length of artery is 100 Pa, the radius is 10 mm, and the flow is laminar. The average speed of the blood is 15 mm/s. (a) What is the net force on the blood in this section of artery? (b) What is the power expended maintaining the flow?

Exercise:

Problem:

A small artery has a length of 1.1×10^{-3} m and a radius of 2.5×10^{-5} m. If the pressure drop across the artery is 1.3 kPa, what is the flow rate through the artery? (Assume that the temperature is 37° C.)

Solution:

$8.7 \times 10^{-11}\ \mathrm{m}^3/\mathrm{s}$

Exercise:

Problem:

Fluid originally flows through a tube at a rate of 100 cm^3/s. To illustrate the sensitivity of flow rate to various factors, calculate the new flow rate for the following changes with all other factors remaining the same as in the original conditions. (a) Pressure difference increases by a factor of 1.50. (b) A new fluid with 3.00 times greater viscosity is substituted. (c) The tube is replaced by one having 4.00 times the length. (d) Another tube is used with a radius 0.100 times the original. (e) Yet another tube is substituted with a radius 0.100 times the original and half the length, *and* the pressure difference is increased by a factor of 1.50.

Exercise:

Problem:

The arterioles (small arteries) leading to an organ, constrict in order to decrease flow to the organ. To shut down an organ, blood flow is reduced naturally to 1.00% of its original value. By what factor did the radii of the arterioles constrict? Penguins do this when they stand on ice to reduce the blood flow to their feet.

Solution:

0.316

Exercise:

Problem:

Angioplasty is a technique in which arteries partially blocked with plaque are dilated to increase blood flow. By what factor must the radius of an artery be increased in order to increase blood flow by a factor of 10?

Exercise:

Problem:

(a) Suppose a blood vessel’s radius is decreased to 90.0% of its original value by plaque deposits and the body compensates by increasing the pressure difference along the vessel to keep the flow rate constant. By what factor must the pressure difference increase? (b) If turbulence is created by the obstruction, what additional effect would it have on the flow rate?

Solution:

(a) 1.52

(b) Turbulence will decrease the flow rate of the blood, which would require an even larger increase in the pressure difference, leading to higher blood pressure.

Exercise:

Problem:

A spherical particle falling at a terminal speed in a liquid must have the gravitational force balanced by the drag force and the buoyant force. The buoyant force is equal to the weight of the displaced fluid, while the drag force is assumed to be given by Stokes Law, $F_s = 6\pi r\eta v$. Show that the terminal speed is given by

Equation:

$$v = \frac{2R^2 g}{9\eta}(\rho_{\text{s}} - \rho_1),$$

where R is the radius of the sphere, ρ_{s} is its density, and ρ_1 is the density of the fluid and η the coefficient of viscosity.

Exercise:

Problem:

Using the equation of the previous problem, find the viscosity of motor oil in which a steel ball of radius 0.8 mm falls with a terminal speed of 4.32 cm/s. The densities of the ball and the oil are 7.86 and 0.88 g/mL, respectively.

Solution:
Equation:

$$225\ \text{mPa} \cdot \text{s}$$

Exercise:

Problem:

A skydiver will reach a terminal velocity when the air drag equals their weight. For a skydiver with high speed and a large body, turbulence is a factor. The drag force then is approximately proportional to the square of the velocity. Taking the drag force to be $F_{\text{D}} = \frac{1}{2}\rho\, A v^2$ and setting this equal to the person's weight, find the terminal speed for a person falling "spread eagle." Find both a formula and a number for v_{t}, with assumptions as to size.

Exercise:

Problem:

A layer of oil 1.50 mm thick is placed between two microscope slides. Researchers find that a force of 5.50×10^{-4} N is required to glide one over the other at a speed of 1.00 cm/s when their contact area is $6.00\ \text{cm}^2$. What is the oil's viscosity? What type of oil might it be?

Solution:
Equation:

$$0.138\ \text{Pa} \cdot \text{s},$$

or

Olive oil.

Exercise:

Problem:

(a) Verify that a 19.0% decrease in laminar flow through a tube is caused by a 5.00% decrease in radius, assuming that all other factors remain constant, as stated in the text. (b) What increase in flow is obtained from a 5.00% increase in radius, again assuming all other factors remain constant?

Exercise:

Problem:

[link] dealt with the flow of saline solution in an IV system. (a) Verify that a pressure of $1.62 \times 10^4 \text{ N/m}^2$ is created at a depth of 1.61 m in a saline solution, assuming its density to be that of sea water. (b) Calculate the new flow rate if the height of the saline solution is decreased to 1.50 m. (c) At what height would the direction of flow be reversed? (This reversal can be a problem when patients stand up.)

Solution:

(a) $1.62 \times 10^4 \text{ N/m}^2$

(b) $0.111 \text{ cm}^3/\text{s}$

(c)10.6 cm

Exercise:

Problem:

When physicians diagnose arterial blockages, they quote the reduction in flow rate. If the flow rate in an artery has been reduced to 10.0% of its normal value by a blood clot and the average pressure difference has increased by 20.0%, by what factor has the clot reduced the radius of the artery?

Exercise:

Problem:

During a marathon race, a runner’s blood flow increases to 10.0 times her resting rate. Her blood’s viscosity has dropped to 95.0% of its normal value, and the blood pressure difference across the circulatory system has increased by 50.0%. By what factor has the average radii of her blood vessels increased?

Solution:

1.59

Exercise:

Problem:

Water supplied to a house by a water main has a pressure of $3.00 \times 10^5\ \mathrm{N/m^2}$ early on a summer day when neighborhood use is low. This pressure produces a flow of 20.0 L/min through a garden hose. Later in the day, pressure at the exit of the water main and entrance to the house drops, and a flow of only 8.00 L/min is obtained through the same hose. (a) What pressure is now being supplied to the house, assuming resistance is constant? (b) By what factor did the flow rate in the water main increase in order to cause this decrease in delivered pressure? The pressure at the entrance of the water main is $5.00 \times 10^5\ \mathrm{N/m^2}$, and the original flow rate was 200 L/min. (c) How many more users are there, assuming each would consume 20.0 L/min in the morning?

Exercise:

Problem:

An oil gusher shoots crude oil 25.0 m into the air through a pipe with a 0.100-m diameter. Neglecting air resistance but not the resistance of the pipe, and assuming laminar flow, calculate the gauge pressure at the entrance of the 50.0-m-long vertical pipe. Take the density of the oil to be 900 $\mathrm{kg/m^3}$ and its viscosity to be 1.00 $(\mathrm{N/m^2}) \cdot \mathrm{s}$ (or 1.00 $\mathrm{Pa \cdot s}$). Note that you must take into account the pressure due to the 50.0-m column of oil in the pipe.

Solution:

$2.95 \times 10^6\ \mathrm{N/m^2}$(gauge pressure)

Exercise:

Problem:

Concrete is pumped from a cement mixer to the place it is being laid, instead of being carried in wheelbarrows. The flow rate is 200.0 L/min through a 50.0-m-long, 8.00-cm-diameter hose, and the pressure at the pump is $8.00 \times 10^6\ \mathrm{N/m^2}$. (a) Calculate the resistance of the hose. (b) What is the viscosity of the concrete, assuming the flow is laminar? (c) How much power is being supplied, assuming the point of use is at the same level as the pump? You may neglect the power supplied to increase the concrete’s velocity.

Exercise:

Problem: Construct Your Own Problem

Consider a coronary artery constricted by arteriosclerosis. Construct a problem in which you calculate the amount by which the diameter of the artery is decreased, based on an assessment of the decrease in flow rate.

Exercise:

Problem:

Consider a river that spreads out in a delta region on its way to the sea. Construct a problem in which you calculate the average speed at which water moves in the delta region, based on the speed at which it was moving up river. Among the things to consider are the size and flow rate of the river before it spreads out and its size once it has spread out. You can construct the problem for the river spreading out into one large river or into multiple smaller rivers.

Glossary

laminar
: a type of fluid flow in which layers do not mix

turbulence
: fluid flow in which layers mix together via eddies and swirls

viscosity
: the friction in a fluid, defined in terms of the friction between layers

Poiseuille's law for resistance
: the resistance to laminar flow of an incompressible fluid in a tube: $R = 8\eta l/\pi r^4$

Poiseuille's law
: the rate of laminar flow of an incompressible fluid in a tube: $Q = (P_2 - P_1)\pi r^4/8\eta l$

Bernoulli’s Equation

- Explain the terms in Bernoulli’s equation.
- Explain how Bernoulli’s equation is related to conservation of energy.
- Explain how to derive Bernoulli’s principle from Bernoulli’s equation.
- Calculate with Bernoulli’s principle.
- List some applications of Bernoulli’s principle.

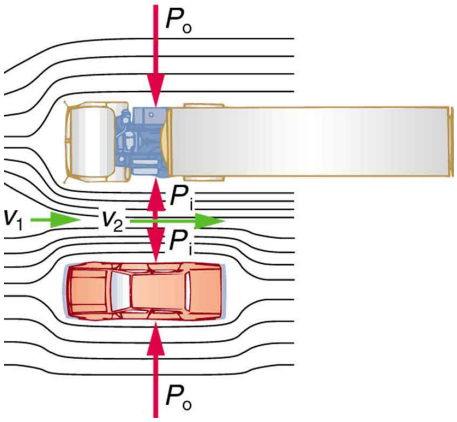
When a fluid flows into a narrower channel, its speed increases. That means its kinetic energy also increases. Where does that change in kinetic energy come from? The increased kinetic energy comes from the net work done on the fluid to push it into the channel and the work done on the fluid by the gravitational force, if the fluid changes vertical position. Recall the work-energy theorem,

Equation:

$$W_{\text{net}} = \frac{1}{2}\text{mv}^2 - \frac{1}{2}\text{mv}_0^2.$$

There is a pressure difference when the channel narrows. This pressure difference results in a net force on the fluid: recall that pressure times area equals force. The net work done increases the fluid’s kinetic energy. As a result, the *pressure will drop in a rapidly-moving fluid*, whether or not the fluid is confined to a tube.

There are a number of common examples of pressure dropping in rapidly-moving fluids. Shower curtains have a disagreeable habit of bulging into the shower stall when the shower is on. The high-velocity stream of water and air creates a region of lower pressure inside the shower, and standard atmospheric pressure on the other side. The pressure difference results in a net force inward pushing the curtain in. You may also have noticed that when passing a truck on the highway, your car tends to veer toward it. The reason is the same—the high velocity of the air between the car and the truck creates a region of lower pressure, and the vehicles are pushed together by greater pressure on the outside. (See [link].) This effect was observed as far back as the mid-1800s, when it was found that trains passing in opposite directions tipped precariously toward one another.



An overhead view of a car passing a truck on a highway. Air passing between the vehicles flows in a narrower channel and must increase its speed (v_2 is greater than v_1), causing the pressure between them to drop (P_i is less than P_o). Greater pressure on the outside pushes the car and truck together.

Note:
Making Connections: Take-Home Investigation with a Sheet of Paper
Hold the short edge of a sheet of paper parallel to your mouth with one hand on each side of your mouth. The page should slant downward over your hands. Blow over the top of the page. Describe what happens and explain the reason for this behavior.

Bernoulli’s Equation

The relationship between pressure and velocity in fluids is described quantitatively by **Bernoulli’s equation**, named after its discoverer, the Swiss scientist Daniel Bernoulli (1700–1782). Bernoulli’s equation states that for an incompressible, frictionless fluid, the following sum is constant:
Equation:

$$P + \frac{1}{2}\rho v^2 + \rho \mathrm{gh} = \mathrm{constant},$$

where P is the absolute pressure, ρ is the fluid density, v is the velocity of the fluid, h is the height above some reference point, and g is the acceleration due to gravity. If we follow a small volume of fluid along its path, various quantities in the sum may change, but the total remains constant. Let the subscripts 1 and 2 refer to any two points along the path that the bit of fluid follows; Bernoulli’s equation becomes
Equation:

$$P_1 + \frac{1}{2}\rho v_1^2 + \rho\, g h_1 = P_2 + \frac{1}{2}\rho v_2^2 + \rho\, g h_2\,.$$

Bernoulli’s equation is a form of the conservation of energy principle. Note that the second and third terms are the kinetic and potential energy with m replaced by ρ. In fact, each term in the equation has units of energy per unit volume. We can prove this for the second term by substituting $\rho = m/V$ into it and gathering terms:
Equation:

$$\frac{1}{2}\rho v^2 = \frac{\frac{1}{2}\mathrm{mv}^2}{V} = \frac{\mathrm{KE}}{V}.$$

So $\frac{1}{2}\rho v^2$ is the kinetic energy per unit volume. Making the same substitution into the third term in the equation, we find
Equation:

$$\rho gh = \frac{mgh}{V} = \frac{\text{PE}_\text{g}}{V},$$

so ρgh is the gravitational potential energy per unit volume. Note that pressure P has units of energy per unit volume, too. Since $P = F/A$, its units are N/m^2. If we multiply these by m/m, we obtain $\text{N} \cdot \text{m}/\text{m}^3 = \text{J}/\text{m}^3$, or energy per unit volume. Bernoulli's equation is, in fact, just a convenient statement of conservation of energy for an incompressible fluid in the absence of friction.

Note:
Making Connections: Conservation of Energy
Conservation of energy applied to fluid flow produces Bernoulli's equation. The net work done by the fluid's pressure results in changes in the fluid's KE and PE_g per unit volume. If other forms of energy are involved in fluid flow, Bernoulli's equation can be modified to take these forms into account. Such forms of energy include thermal energy dissipated because of fluid viscosity.

The general form of Bernoulli's equation has three terms in it, and it is broadly applicable. To understand it better, we will look at a number of specific situations that simplify and illustrate its use and meaning.

Bernoulli's Equation for Static Fluids

Let us first consider the very simple situation where the fluid is static—that is, $v_1 = v_2 = 0$. Bernoulli's equation in that case is
Equation:

$$P_1 + \rho gh_1 = P_2 + \rho gh_2\,.$$

We can further simplify the equation by taking $h_2 = 0$ (we can always choose some height to be zero, just as we often have done for other situations involving the gravitational force, and take all other heights to be relative to this). In that case, we get
Equation:

$$P_2 = P_1 + \rho\, g h_1\,.$$

This equation tells us that, in static fluids, pressure increases with depth. As we go from point 1 to point 2 in the fluid, the depth increases by h_1, and consequently, P_2 is greater than P_1 by an amount $\rho\, g h_1$. In the very simplest case, P_1 is zero at the top of the fluid, and we get the familiar relationship $P = \rho\, g h$. (Recall that $P = \rho g h$ and $\Delta \mathrm{PE_g} = \mathrm{mgh}$.) Bernoulli's equation includes the fact that the pressure due to the weight of a fluid is ρgh. Although we introduce Bernoulli's equation for fluid flow, it includes much of what we studied for static fluids in the preceding chapter.

Bernoulli's Principle—Bernoulli's Equation at Constant Depth

Another important situation is one in which the fluid moves but its depth is constant—that is, $h_1 = h_2$. Under that condition, Bernoulli's equation becomes
Equation:

$$P_1 + \frac{1}{2}\rho v_1^2 = P_2 + \frac{1}{2}\rho v_2^2.$$

Situations in which fluid flows at a constant depth are so important that this equation is often called **Bernoulli's principle**. It is Bernoulli's equation for fluids at constant depth. (Note again that this applies to a small volume of fluid as we follow it along its path.) As we have just discussed, pressure drops as speed increases in a moving fluid. We can see this from Bernoulli's principle. For example, if v_2 is greater than v_1 in the equation, then P_2 must be less than P_1 for the equality to hold.

Example:
Calculating Pressure: Pressure Drops as a Fluid Speeds Up
In [link], we found that the speed of water in a hose increased from 1.96 m/s to 25.5 m/s going from the hose to the nozzle. Calculate the pressure in the hose, given that the absolute pressure in the nozzle is $1.01 \times 10^5 \text{ N/m}^2$ (atmospheric, as it must be) and assuming level, frictionless flow.
Strategy
Level flow means constant depth, so Bernoulli's principle applies. We use the subscript 1 for values in the hose and 2 for those in the nozzle. We are thus asked to find P_1.
Solution
Solving Bernoulli's principle for P_1 yields
Equation:

$$P_1 = P_2 + \frac{1}{2}\rho v_2^2 - \frac{1}{2}\rho v_1^2 = P_2 + \frac{1}{2}\rho(v_2^2 - v_1^2).$$

Substituting known values,
Equation:

$$\begin{aligned} P_1 &= 1.01 \times 10^5 \text{ N/m}^2 \\ &\quad + \tfrac{1}{2}(10^3 \text{ kg/m}^3)\left[(25.5 \text{ m/s})^2 - (1.96 \text{ m/s})^2\right] \\ &= 4.24 \times 10^5 \text{ N/m}^2. \end{aligned}$$

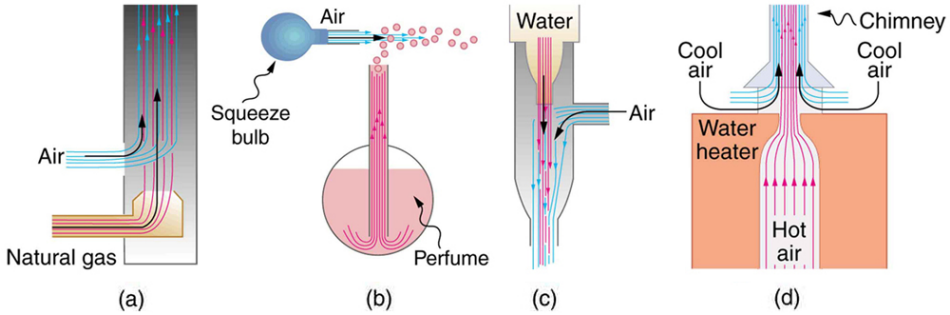
Discussion
This absolute pressure in the hose is greater than in the nozzle, as expected since v is greater in the nozzle. The pressure P_2 in the nozzle must be atmospheric since it emerges into the atmosphere without other changes in conditions.

Applications of Bernoulli's Principle

There are a number of devices and situations in which fluid flows at a constant height and, thus, can be analyzed with Bernoulli's principle.

Entrainment

People have long put the Bernoulli principle to work by using reduced pressure in high-velocity fluids to move things about. With a higher pressure on the outside, the high-velocity fluid forces other fluids into the stream. This process is called *entrainment*. Entrainment devices have been in use since ancient times, particularly as pumps to raise water small heights, as in draining swamps, fields, or other low-lying areas. Some other devices that use the concept of entrainment are shown in [link].



Examples of entrainment devices that use increased fluid speed to create low pressures, which then entrain one fluid into another. (a) A Bunsen burner uses an adjustable gas nozzle, entraining air for proper combustion. (b) An atomizer uses a squeeze bulb to create a jet of air that entrains drops of perfume. Paint sprayers and carburetors use very similar techniques to move their respective liquids. (c) A common aspirator uses a high-speed stream of water to create a region of lower pressure. Aspirators may be used as suction pumps in dental and surgical situations or for draining a flooded basement or producing a reduced pressure in a vessel. (d) The chimney of a water heater is designed to entrain air into the pipe leading through the ceiling.

Wings and Sails

The airplane wing is a beautiful example of Bernoulli's principle in action. [link](a) shows the characteristic shape of a wing. The wing is tilted upward at a small angle and the upper surface is longer, causing air to flow faster over it. The pressure on top of the wing is therefore reduced, creating a net upward force or lift. (Wings can also gain lift by pushing air downward, utilizing the conservation of momentum principle. The deflected air molecules result in an upward force on the wing — Newton's third law.) Sails also have the characteristic shape of a wing. (See [link](b).) The pressure on the front side of the sail, P_{front}, is lower than the pressure on the back of the sail, P_{back}. This results in a forward force and even allows you to sail into the wind.

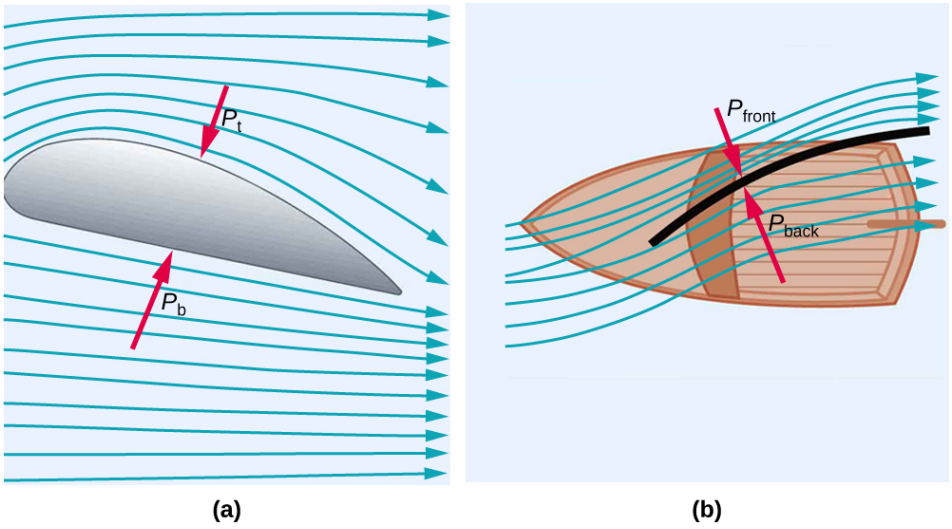
Note:
Making Connections: Take-Home Investigation with Two Strips of Paper
For a good illustration of Bernoulli's principle, make two strips of paper, each about 15 cm long and 4 cm wide. Hold the small end of one strip up to your lips and let it drape over your finger. Blow across the paper. What happens? Now hold two strips of paper up to your lips, separated by your fingers. Blow between the strips. What happens?

Velocity measurement

[link] shows two devices that measure fluid velocity based on Bernoulli's principle. The manometer in [link](a) is connected to two tubes that are small enough not to appreciably disturb the flow. The tube facing the oncoming fluid creates a dead spot having zero velocity ($v_1 = 0$) in front of it, while fluid passing the other tube has velocity v_2. This means that Bernoulli's principle as stated in $P_1 + \frac{1}{2}\rho v_1^2 = P_2 + \frac{1}{2}\rho v_2^2$ becomes

Equation:

$$P_1 = P_2 + \frac{1}{2}\rho v_2^2.$$



(a) The Bernoulli principle helps explain lift generated by a wing. (b) Sails use the same technique to generate part of their thrust.

Thus pressure P_2 over the second opening is reduced by $\frac{1}{2}\rho v_2^2$, and so the fluid in the manometer rises by h on the side connected to the second opening, where

Equation:

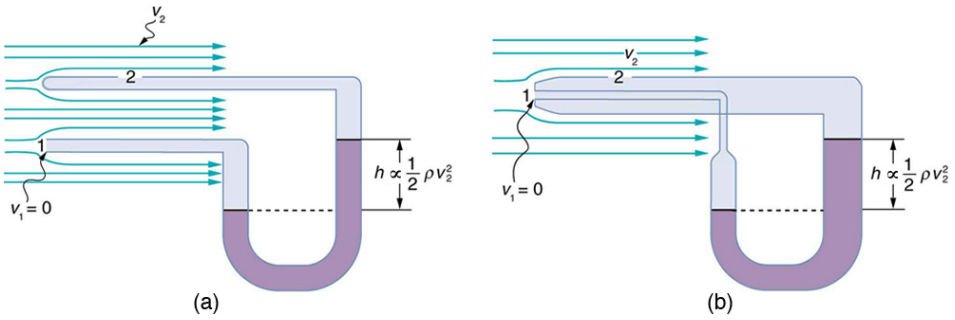
$$h \propto \frac{1}{2}\rho v_2^2.$$

(Recall that the symbol $\propto$ means "proportional to.") Solving for v_2, we see that

Equation:

$$v_2 \propto \sqrt{h}.$$

[link](b) shows a version of this device that is in common use for measuring various fluid velocities; such devices are frequently used as air speed indicators in aircraft.



Measurement of fluid speed based on Bernoulli's principle. (a) A manometer is connected to two tubes that are close together and small enough not to disturb the flow. Tube 1 is open at the end facing the flow. A dead spot having zero speed is created there. Tube 2 has an opening on the side, and so the fluid has a speed v across the opening; thus, pressure there drops. The difference in pressure at the manometer is $\frac{1}{2}\rho v_2^2$, and so h is proportional to $\frac{1}{2}\rho v_2^2$. (b) This type of velocity measuring device is a Prandtl tube, also known as a pitot tube.

Summary

- Bernoulli's equation states that the sum on each side of the following equation is constant, or the same at any two points in an incompressible frictionless fluid:
  Equation:

$$P_1 + \frac{1}{2}\rho v_1^2 + \rho\, g h_1 = P_2 + \frac{1}{2}\rho v_2^2 + \rho \mathrm{g} \mathrm{h}_2.$$

- Bernoulli’s principle is Bernoulli’s equation applied to situations in which depth is constant. The terms involving depth (or height *h*) subtract out, yielding
  Equation:

$$P_1 + \frac{1}{2}\rho v_1^2 = P_2 + \frac{1}{2}\rho v_2^2.$$

- Bernoulli’s principle has many applications, including entrainment, wings and sails, and velocity measurement.

Conceptual Questions

Exercise:

Problem:

You can squirt water a considerably greater distance by placing your thumb over the end of a garden hose and then releasing, than by leaving it completely uncovered. Explain how this works.

Exercise:

Problem:

Water is shot nearly vertically upward in a decorative fountain and the stream is observed to broaden as it rises. Conversely, a stream of water falling straight down from a faucet narrows. Explain why, and discuss whether surface tension enhances or reduces the effect in each case.

Exercise:

Problem:

Look back to [link]. Answer the following two questions. Why is P_{o} less than atmospheric? Why is P_{o} greater than P_{i}?

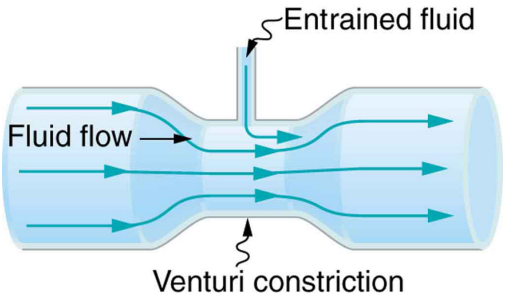
Exercise:

Problem: Give an example of entrainment not mentioned in the text.

Exercise:

Problem:

Many entrainment devices have a constriction, called a Venturi, such as shown in [link]. How does this bolster entrainment?



A tube with a narrow segment designed to enhance entrainment is called a Venturi. These are very commonly used in carburetors and aspirators.

Exercise:

Problem:

Some chimney pipes have a T-shape, with a crosspiece on top that helps draw up gases whenever there is even a slight breeze. Explain how this works in terms of Bernoulli's principle.

Exercise:

Problem:

Is there a limit to the height to which an entrainment device can raise a fluid? Explain your answer.

Exercise:

Problem:

Why is it preferable for airplanes to take off into the wind rather than with the wind?

Exercise:

Problem:

Roofs are sometimes pushed off vertically during a tropical cyclone, and buildings sometimes explode outward when hit by a tornado. Use Bernoulli's principle to explain these phenomena.

Exercise:

Problem: Why does a sailboat need a keel?

Exercise:

Problem:

It is dangerous to stand close to railroad tracks when a rapidly moving commuter train passes. Explain why atmospheric pressure would push you toward the moving train.

Exercise:

Problem:

Water pressure inside a hose nozzle can be less than atmospheric pressure due to the Bernoulli effect. Explain in terms of energy how the water can emerge from the nozzle against the opposing atmospheric pressure.

Exercise:

Problem:

A perfume bottle or atomizer sprays a fluid that is in the bottle. ([link].) How does the fluid rise up in the vertical tube in the bottle?



Atomizer: perfume bottle with tube to carry perfume up through the bottle. (credit: Antonia Foy, Flickr)

Exercise:

Problem:

If you lower the window on a car while moving, an empty plastic bag can sometimes fly out the window. Why does this happen?

Problems & Exercises

Exercise:

Problem: Verify that pressure has units of energy per unit volume.

Solution:

$$\begin{aligned} P &= \frac{\text{Force}}{\text{Area}}, \\ (P)_{\text{units}} &= \text{N/m}^2 = \text{N} \cdot \text{m/m}^3 = \text{J/m}^3 \\ &= \text{energy/volume} \end{aligned}$$

Exercise:

Problem:

Suppose you have a wind speed gauge like the pitot tube shown in [link](b). By what factor must wind speed increase to double the value of h in the manometer? Is this independent of the moving fluid and the fluid in the manometer?

Exercise:

Problem:

If the pressure reading of your pitot tube is 15.0 mm Hg at a speed of 200 km/h, what will it be at 700 km/h at the same altitude?

Solution:

184 mm Hg

Exercise:

Problem:

Calculate the maximum height to which water could be squirted with the hose in [link] example if it: (a) Emerges from the nozzle. (b) Emerges with the nozzle removed, assuming the same flow rate.

Exercise:

Problem:

Every few years, winds in Boulder, Colorado, attain sustained speeds of 45.0 m/s (about 100 mi/h) when the jet stream descends during early spring. Approximately what is the force due to the Bernoulli effect on a roof having an area of 220 m^2? Typical air density in Boulder is 1.14 $\mathrm{kg/m}^3$, and the corresponding atmospheric pressure is $8.89 \times 10^4\ \mathrm{N/m}^2$. (Bernoulli's principle as stated in the text assumes laminar flow. Using the principle here produces only an approximate result, because there is significant turbulence.)

Solution:

$2.54 \times 10^5\ \mathrm{N}$

Exercise:

Problem:

(a) Calculate the approximate force on a square meter of sail, given the horizontal velocity of the wind is 6.00 m/s parallel to its front surface and 3.50 m/s along its back surface. Take the density of air to be 1.29 $\mathrm{kg/m}^3$. (The calculation, based on Bernoulli's principle, is approximate due to the effects of turbulence.) (b) Discuss whether this force is great enough to be effective for propelling a sailboat.

Exercise:

Problem:

(a) What is the pressure drop due to the Bernoulli effect as water goes into a 3.00-cm-diameter nozzle from a 9.00-cm-diameter fire hose while carrying a flow of 40.0 L/s? (b) To what maximum height above the nozzle can this water rise? (The actual height will be significantly smaller due to air resistance.)

Solution:

(a) $1.58 \times 10^6\ \mathrm{N/m}^2$

(b) 163 m

Exercise:

Problem:

(a) Using Bernoulli’s equation, show that the measured fluid speed v for a pitot tube, like the one in [link](b), is given by

Equation:

$$v = \left(\frac{2\rho\prime gh}{\rho}\right)^{1/2},$$

where h is the height of the manometer fluid, $\rho\prime$ is the density of the manometer fluid, ρ is the density of the moving fluid, and g is the acceleration due to gravity. (Note that v is indeed proportional to the square root of h, as stated in the text.) (b) Calculate v for moving air if a mercury manometer’s h is 0.200 m.

Glossary

Bernoulli’s equation
: the equation resulting from applying conservation of energy to an incompressible frictionless fluid: $P + 1/2pv^2 + pgh$ = constant , through the fluid

Bernoulli’s principle
: Bernoulli’s equation applied at constant depth: $P_1 + 1/2pv_1^2 = P_2 + 1/2pv_2^2$

The Most General Applications of Bernoulli’s Equation

- Calculate using Torricelli’s theorem.
- Calculate power in fluid flow.

Torricelli’s Theorem

[link] shows water gushing from a large tube through a dam. What is its speed as it emerges? Interestingly, if resistance is negligible, the speed is just what it would be if the water fell a distance h from the surface of the reservoir; the water’s speed is independent of the size of the opening. Let us check this out. Bernoulli’s equation must be used since the depth is not constant. We consider water flowing from the surface (point 1) to the tube’s outlet (point 2). Bernoulli’s equation as stated in previously is
Equation:

$$P_1 + \frac{1}{2}\rho v_1^2 + \rho g h_1 = P_2 + \frac{1}{2}\rho v_2^2 + \rho g h_2.$$

Both P_1 and P_2 equal atmospheric pressure (P_1 is atmospheric pressure because it is the pressure at the top of the reservoir. P_2 must be atmospheric pressure, since the emerging water is surrounded by the atmosphere and cannot have a pressure different from atmospheric pressure.) and subtract out of the equation, leaving
Equation:

$$\frac{1}{2}\rho v_1^2 + \rho g h_1 = \frac{1}{2}\rho v_2^2 + \rho g h_2.$$

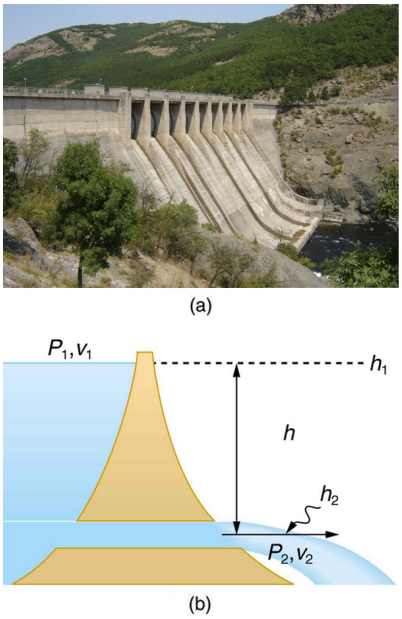
Solving this equation for v_2^2, noting that the density ρ cancels (because the fluid is incompressible), yields
Equation:

$$v_2^2 = v_1^2 + 2g(h_1 - h_2).$$

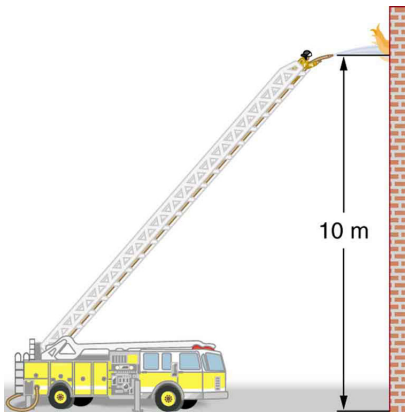
We let $h = h_1 - h_2$; the equation then becomes
Equation:

$$v_2^2 = v_1^2 + 2\mathrm{gh}$$

where h is the height dropped by the water. This is simply a kinematic equation for any object falling a distance h with negligible resistance. In fluids, this last equation is called *Torricelli’s theorem*. Note that the result is independent of the velocity’s direction, just as we found when applying conservation of energy to falling objects.



(a) Water gushes from the base of the Studen Kladenetz dam in Bulgaria. (credit: Kiril Kapustin; http://www.ImagesFromBulgaria.com) (b) In the absence of significant resistance, water flows from the reservoir with the same speed it would have if it fell the distance h without friction. This is an example of Torricelli's theorem.



Pressure in the nozzle of this fire hose is less than at ground level for two reasons: the water has to go uphill to get to the nozzle, and speed increases in the nozzle. In spite of its

lowered pressure, the water can exert a large force on anything it strikes, by virtue of its kinetic energy. Pressure in the water stream becomes equal to atmospheric pressure once it emerges into the air.

All preceding applications of Bernoulli's equation involved simplifying conditions, such as constant height or constant pressure. The next example is a more general application of Bernoulli's equation in which pressure, velocity, and height all change. (See [link].)

Example:
Calculating Pressure: A Fire Hose Nozzle
Fire hoses used in major structure fires have inside diameters of 6.40 cm. Suppose such a hose carries a flow of 40.0 L/s starting at a gauge pressure of $1.62 \times 10^6\ \mathrm{N/m^2}$. The hose goes 10.0 m up a ladder to a nozzle having an inside diameter of 3.00 cm. Assuming negligible resistance, what is the pressure in the nozzle?
Strategy
Here we must use Bernoulli's equation to solve for the pressure, since depth is not constant.
Solution
Bernoulli's equation states
Equation:

$$P_1 + \frac{1}{2}\rho v_1^2 + \rho g h_1 = P_2 + \frac{1}{2}\rho v_2^2 + \rho g h_2,$$

where the subscripts 1 and 2 refer to the initial conditions at ground level and the final conditions inside the nozzle, respectively. We must first find the speeds v_1 and v_2. Since $Q = A_1 v_1$, we get
Equation:

$$v_1 = \frac{Q}{A_1} = \frac{40.0 \times 10^{-3}\ \mathrm{m^3/s}}{\pi(3.20 \times 10^{-2}\ \mathrm{m})^2} = 12.4\ \mathrm{m/s}.$$

Similarly, we find
Equation:

$$v_2 = 56.6\ \mathrm{m/s}.$$

(This rather large speed is helpful in reaching the fire.) Now, taking h_1 to be zero, we solve Bernoulli's equation for P_2:
Equation:

$$P_2 = P_1 + \frac{1}{2}\rho\left(v_1^2 - v_2^2\right) - \rho g h_2.$$

Substituting known values yields
Equation:

$$P_2 = 1.62 \times 10^6\ \mathrm{N/m^2} + \frac{1}{2}(1000\ \mathrm{kg/m^3})\left[(12.4\ \mathrm{m/s})^2 - (56.6\ \mathrm{m/s})^2\right] - (1000\ \mathrm{kg/m^3})(9.80\ \mathrm{m/s^2})(10.0\ \mathrm{m}$$

Discussion

This value is a gauge pressure, since the initial pressure was given as a gauge pressure. Thus the nozzle pressure equals atmospheric pressure, as it must because the water exits into the atmosphere without changes in its conditions.

Power in Fluid Flow

Power is the *rate* at which work is done or energy in any form is used or supplied. To see the relationship of power to fluid flow, consider Bernoulli's equation:
Equation:

$$P + \frac{1}{2}\rho v^2 + \rho \text{gh} = \text{constant}.$$

All three terms have units of energy per unit volume, as discussed in the previous section. Now, considering units, if we multiply energy per unit volume by flow rate (volume per unit time), we get units of power. That is, $(E/V)(V/t) = E/t$. This means that if we multiply Bernoulli's equation by flow rate Q, we get power. In equation form, this is
Equation:

$$\left(P + \frac{1}{2}\rho v^2 + \rho \text{gh}\right)Q = \text{power}.$$

Each term has a clear physical meaning. For example, PQ is the power supplied to a fluid, perhaps by a pump, to give it its pressure P. Similarly, $\frac{1}{2}\rho v^2 Q$ is the power supplied to a fluid to give it its kinetic energy. And ρghQ is the power going to gravitational potential energy.

Note:
Making Connections: Power
Power is defined as the rate of energy transferred, or E/t. Fluid flow involves several types of power. Each type of power is identified with a specific type of energy being expended or changed in form.

Example:
Calculating Power in a Moving Fluid
Suppose the fire hose in the previous example is fed by a pump that receives water through a hose with a 6.40-cm diameter coming from a hydrant with a pressure of $0.700 \times 10^6\ \text{N/m}^2$. What power does the pump supply to the water?
Strategy
Here we must consider energy forms as well as how they relate to fluid flow. Since the input and output hoses have the same diameters and are at the same height, the pump does not change the speed of the water nor its height, and so the water's kinetic energy and gravitational potential energy are unchanged. That means the pump only supplies power to increase water pressure by $0.92 \times 10^6\ \text{N/m}^2$ (from $0.700 \times 10^6\ \text{N/m}^2$ to $1.62 \times 10^6\ \text{N/m}^2$).
Solution
As discussed above, the power associated with pressure is
Equation:

$$\begin{aligned} \text{power} &= \text{PQ} \\ &= \left(0.920 \times 10^6\ \text{N/m}^2\right)\left(40.0 \times 10^{-3}\ \text{m}^3/\text{s}\right). \\ &= 3.68 \times 10^4\ \text{W} = 36.8\ \text{kW} \end{aligned}$$

Discussion
Such a substantial amount of power requires a large pump, such as is found on some fire trucks. (This kilowatt value converts to about 50 hp.) The pump in this example increases only the water's pressure. If a pump—such as the heart—directly increases velocity and height as well as pressure, we would have to calculate all three terms to find the power it supplies.

Summary

- Power in fluid flow is given by the equation $\left(P_1 + \frac{1}{2}\rho v^2 + \rho \text{gh}\right)Q = \text{power}$, where the first term is power associated with pressure, the second is power associated with velocity, and the third is power associated with height.

Conceptual Questions

Exercise:

Problem:

Based on Bernoulli's equation, what are three forms of energy in a fluid? (Note that these forms are conservative, unlike heat transfer and other dissipative forms not included in Bernoulli's equation.)

Exercise:

Problem:

Water that has emerged from a hose into the atmosphere has a gauge pressure of zero. Why? When you put your hand in front of the emerging stream you feel a force, yet the water's gauge pressure is zero. Explain where the force comes from in terms of energy.

Exercise:

Problem:

The old rubber boot shown in [link] has two leaks. To what maximum height can the water squirt from Leak 1? How does the velocity of water emerging from Leak 2 differ from that of leak 1? Explain your responses in terms of energy.



Water emerges from two leaks in an old boot.

Exercise:

Problem:

Water pressure inside a hose nozzle can be less than atmospheric pressure due to the Bernoulli effect. Explain in terms of energy how the water can emerge from the nozzle against the opposing atmospheric pressure.

Problems & Exercises

Exercise:

Problem:

Hoover Dam on the Colorado River is the highest dam in the United States at 221 m, with an output of 1300 MW. The dam generates electricity with water taken from a depth of 150 m and an average flow rate of $650\ \mathrm{m^3/s}$. (a) Calculate the power in this flow. (b) What is the ratio of this power to the facility's average of 680 MW?

Solution:

(a) $9.56 \times 10^8\ \mathrm{W}$

(b) 1.4

Exercise:

Problem:

A frequently quoted rule of thumb in aircraft design is that wings should produce about 1000 N of lift per square meter of wing. (The fact that a wing has a top and bottom surface does not double its area.) (a) At takeoff, an aircraft travels at 60.0 m/s, so that the air speed relative to the bottom of the wing is 60.0 m/s. Given the sea level density of air to be $1.29\ \mathrm{kg/m^3}$, how fast must it move over the upper surface to create the ideal lift? (b) How fast must air move over the upper surface at a cruising speed of 245 m/s and at an altitude where air density is one-fourth that at sea level? (Note that this is not all of the aircraft's lift—some comes from the body of the plane, some from engine thrust, and so on. Furthermore, Bernoulli's principle gives an approximate answer because flow over the wing creates turbulence.)

Exercise:

Problem:

The left ventricle of a resting adult's heart pumps blood at a flow rate of $83.0\ \mathrm{cm^3/s}$, increasing its pressure by 110 mm Hg, its speed from zero to 30.0 cm/s, and its height by 5.00 cm. (All numbers are averaged over the entire heartbeat.) Calculate the total power output of the left ventricle. Note that most of the power is used to increase blood pressure.

Solution:

1.26 W

Exercise:

Problem:

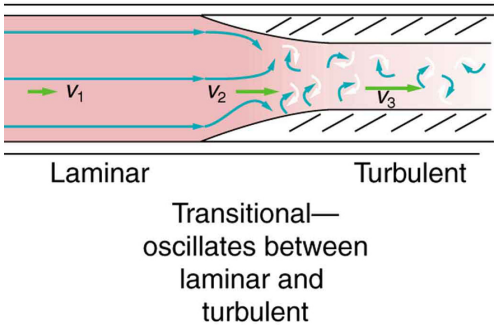
A sump pump (used to drain water from the basement of houses built below the water table) is draining a flooded basement at the rate of 0.750 L/s, with an output pressure of $3.00 \times 10^5\ \mathrm{N/m^2}$. (a) The water enters a hose with a 3.00-cm inside diameter and rises 2.50 m above the pump. What is its pressure at this point? (b) The hose goes over the foundation wall, losing 0.500 m in height, and widens to 4.00 cm in diameter. What is the pressure now? You may neglect frictional losses in both parts of the problem.

The Onset of Turbulence

- Calculate Reynolds number.
- Use the Reynolds number for a system to determine whether it is laminar or turbulent.

Sometimes we can predict if flow will be laminar or turbulent. We know that flow in a very smooth tube or around a smooth, streamlined object will be laminar at low velocity. We also know that at high velocity, even flow in a smooth tube or around a smooth object will experience turbulence. In between, it is more difficult to predict. In fact, at intermediate velocities, flow may oscillate back and forth indefinitely between laminar and turbulent.

An occlusion, or narrowing, of an artery, such as shown in [link], is likely to cause turbulence because of the irregularity of the blockage, as well as the complexity of blood as a fluid. Turbulence in the circulatory system is noisy and can sometimes be detected with a stethoscope, such as when measuring diastolic pressure in the upper arm's partially collapsed brachial artery. These turbulent sounds, at the onset of blood flow when the cuff pressure becomes sufficiently small, are called *Korotkoff sounds*. Aneurysms, or ballooning of arteries, create significant turbulence and can sometimes be detected with a stethoscope. Heart murmurs, consistent with their name, are sounds produced by turbulent flow around damaged and insufficiently closed heart valves. Ultrasound can also be used to detect turbulence as a medical indicator in a process analogous to Doppler-shift radar used to detect storms.



Flow is laminar in the large part of this blood

vessel and turbulent in the part narrowed by plaque, where velocity is high. In the transition region, the flow can oscillate chaotically between laminar and turbulent flow.

An indicator called the **Reynolds number** N_{R} can reveal whether flow is laminar or turbulent. For flow in a tube of uniform diameter, the Reynolds number is defined as

Equation:

$$N_{\mathrm{R}} = \frac{2\rho \mathrm{vr}}{\eta} (\text{flow in tube}),$$

where ρ is the fluid density, v its speed, η its viscosity, and r the tube radius. The Reynolds number is a unitless quantity. Experiments have revealed that N_{R} is related to the onset of turbulence. For N_{R} below about 2000, flow is laminar. For N_{R} above about 3000, flow is turbulent. For values of N_{R} between about 2000 and 3000, flow is unstable—that is, it can be laminar, but small obstructions and surface roughness can make it turbulent, and it may oscillate randomly between being laminar and turbulent. The blood flow through most of the body is a quiet, laminar flow. The exception is in the aorta, where the speed of the blood flow rises above a critical value of 35 m/s and becomes turbulent.

Example:
Is This Flow Laminar or Turbulent?
Calculate the Reynolds number for flow in the needle considered in Example 12.8 to verify the assumption that the flow is laminar. Assume

that the density of the saline solution is $1025\ \mathrm{kg/m^3}$.

Strategy

We have all of the information needed, except the fluid speed v, which can be calculated from $v = Q/A = 1.70\ \mathrm{m/s}$ (verification of this is in this chapter's Problems and Exercises).

Solution

Entering the known values into $N_\mathrm{R} = \frac{2\rho \mathrm{v} r}{\eta}$ gives

Equation:

$$\begin{aligned} N_\mathrm{R} &= \frac{2\rho \mathrm{v} r}{\eta} \\ &= \frac{2(1025\ \mathrm{kg/m^3})(1.70\ \mathrm{m/s})(0.150\times10^{-3}\ \mathrm{m})}{1.00\times10^{-3}\ \mathrm{N\cdot s/m^2}} \\ &= 523. \end{aligned}$$

Discussion

Since N_R is well below 2000, the flow should indeed be laminar.

Note:

Take-Home Experiment: Inhalation

Under the conditions of normal activity, an adult inhales about 1 L of air during each inhalation. With the aid of a watch, determine the time for one of your own inhalations by timing several breaths and dividing the total length by the number of breaths. Calculate the average flow rate Q of air traveling through the trachea during each inhalation.

The topic of chaos has become quite popular over the last few decades. A system is defined to be *chaotic* when its behavior is so sensitive to some factor that it is extremely difficult to predict. The field of *chaos* is the study of chaotic behavior. A good example of chaotic behavior is the flow of a fluid with a Reynolds number between 2000 and 3000. Whether or not the flow is turbulent is difficult, but not impossible, to predict—the difficulty lies in the extremely sensitive dependence on factors like roughness and

obstructions on the nature of the flow. A tiny variation in one factor has an exaggerated (or nonlinear) effect on the flow. Phenomena as disparate as turbulence, the orbit of Pluto, and the onset of irregular heartbeats are chaotic and can be analyzed with similar techniques.

Section Summary

- The Reynolds number N_{R} can reveal whether flow is laminar or turbulent. It is
  Equation:

$$N_{\mathrm{R}} = \frac{2\rho \mathrm{vr}}{\eta}.$$

- For N_{R} below about 2000, flow is laminar. For N_{R} above about 3000, flow is turbulent. For values of N_{R} between 2000 and 3000, it may be either or both.

Conceptual Questions

Exercise:

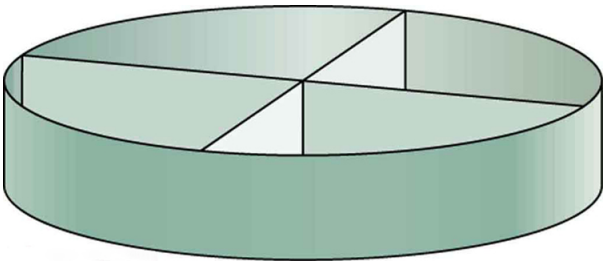
Problem:

Doppler ultrasound can be used to measure the speed of blood in the body. If there is a partial constriction of an artery, where would you expect blood speed to be greatest, at or nearby the constriction? What are the two distinct causes of higher resistance in the constriction?

Exercise:

Problem:

Sink drains often have a device such as that shown in [link] to help speed the flow of water. How does this work?



You will find devices such as this in many drains. They significantly increase flow rate.

Exercise:

Problem:

Some ceiling fans have decorative wicker reeds on their blades. Discuss whether these fans are as quiet and efficient as those with smooth blades.

Problems & Exercises

Exercise:

Problem:

Verify that the flow of oil is laminar (barely) for an oil gusher that shoots crude oil 25.0 m into the air through a pipe with a 0.100-m diameter. The vertical pipe is 50 m long. Take the density of the oil to be 900 $\mathrm{kg/m^3}$ and its viscosity to be 1.00 $(\mathrm{N/m^2}) \cdot \mathrm{s}$ (or 1.00 $\mathrm{Pa \cdot s}$).

Solution:

$N_{\mathrm{R}} = 1.99 \times 10^2 < 2000$

Exercise:

Problem:

Show that the Reynolds number N_{R} is unitless by substituting units for all the quantities in its definition and cancelling.

Exercise:

Problem:

Calculate the Reynolds numbers for the flow of water through (a) a nozzle with a radius of 0.250 cm and (b) a garden hose with a radius of 0.900 cm, when the nozzle is attached to the hose. The flow rate through hose and nozzle is 0.500 L/s. Can the flow in either possibly be laminar?

Solution:

(a) nozzle: 1.27×10^{5} , not laminar

(b) hose: 3.51×10^{4}, not laminar.

Exercise:

Problem:

A fire hose has an inside diameter of 6.40 cm. Suppose such a hose carries a flow of 40.0 L/s starting at a gauge pressure of $1.62 \times 10^{6}\ \mathrm{N/m^{2}}$. The hose goes 10.0 m up a ladder to a nozzle having an inside diameter of 3.00 cm. Calculate the Reynolds numbers for flow in the fire hose and nozzle to show that the flow in each must be turbulent.

Exercise:

Problem:

Concrete is pumped from a cement mixer to the place it is being laid, instead of being carried in wheelbarrows. The flow rate is 200.0 L/min through a 50.0-m-long, 8.00-cm-diameter hose, and the pressure at the pump is $8.00 \times 10^6\ \mathrm{N/m^2}$. Verify that the flow of concrete is laminar taking concrete's viscosity to be $48.0\ (\mathrm{N/m^2}) \cdot \mathrm{s}$, and given its density is $2300\ \mathrm{kg/m^3}$.

Solution:

2.54 << 2000, laminar.

Exercise:

Problem:

At what flow rate might turbulence begin to develop in a water main with a 0.200-m diameter? Assume a $20^{\mathrm{o}}\ \mathrm{C}$ temperature.

Exercise:

Problem:

What is the greatest average speed of blood flow at $37^{\mathrm{o}}\ \mathrm{C}$ in an artery of radius 2.00 mm if the flow is to remain laminar? What is the corresponding flow rate? Take the density of blood to be $1025\ \mathrm{kg/m^3}$.

Solution:

1.02 m/s

$1.28 \times 10^{-2}\ \mathrm{L/s}$

Exercise:

Problem:

In Take-Home Experiment: Inhalation, we measured the average flow rate Q of air traveling through the trachea during each inhalation. Now calculate the average air speed in meters per second through your trachea during each inhalation. The radius of the trachea in adult humans is approximately 10^{-2} m. From the data above, calculate the Reynolds number for the air flow in the trachea during inhalation. Do you expect the air flow to be laminar or turbulent?

Exercise:

Problem:

Gasoline is piped underground from refineries to major users. The flow rate is $3.00 \times 10^{-2}\ \mathrm{m^3/s}$ (about 500 gal/min), the viscosity of gasoline is $1.00 \times 10^{-3}\ (\mathrm{N/m^2}) \cdot \mathrm{s}$, and its density is $680\ \mathrm{kg/m^3}$. (a) What minimum diameter must the pipe have if the Reynolds number is to be less than 2000? (b) What pressure difference must be maintained along each kilometer of the pipe to maintain this flow rate?

Solution:

(a) $\geq 13.0\ \mathrm{m}$

(b) $2.68 \times 10^{-6}\ \mathrm{N/m^2}$

Exercise:

Problem:

Assuming that blood is an ideal fluid, calculate the critical flow rate at which turbulence is a certainty in the aorta. Take the diameter of the aorta to be 2.50 cm. (Turbulence will actually occur at lower average flow rates, because blood is not an ideal fluid. Furthermore, since blood flow pulses, turbulence may occur during only the high-velocity part of each heartbeat.)

Exercise:

Problem: **Unreasonable Results**

A fairly large garden hose has an internal radius of 0.600 cm and a length of 23.0 m. The nozzleless horizontal hose is attached to a faucet, and it delivers 50.0 L/s. (a) What water pressure is supplied by the faucet? (b) What is unreasonable about this pressure? (c) What is unreasonable about the premise? (d) What is the Reynolds number for the given flow? (Take the viscosity of water as $1.005 \times 10^{-3}\ \left(\mathrm{N/m^2}\right) \cdot \mathrm{s}$.)

Solution:

(a) 23.7 atm or $344\ \mathrm{lb/in^2}$

(b) The pressure is much too high.

(c) The assumed flow rate is very high for a garden hose.

(d) $5.27 \times 10^6 > > 3000$, turbulent, contrary to the assumption of laminar flow when using this equation.

Glossary

Reynolds number
: a dimensionless parameter that can reveal whether a particular flow is laminar or turbulent

Motion of an Object in a Viscous Fluid

- Calculate the Reynolds number for an object moving through a fluid.
- Explain whether the Reynolds number indicates laminar or turbulent flow.
- Describe the conditions under which an object has a terminal speed.

A moving object in a viscous fluid is equivalent to a stationary object in a flowing fluid stream. (For example, when you ride a bicycle at 10 m/s in still air, you feel the air in your face exactly as if you were stationary in a 10-m/s wind.) Flow of the stationary fluid around a moving object may be laminar, turbulent, or a combination of the two. Just as with flow in tubes, it is possible to predict when a moving object creates turbulence. We use another form of the Reynolds number $N\prime_{\mathrm{R}}$, defined for an object moving in a fluid to be

Equation:

$$N\prime_{\mathrm{R}} = \frac{\rho \mathrm{vL}}{\eta} \text{(object in fluid)},$$

where L is a characteristic length of the object (a sphere's diameter, for example), ρ the fluid density, η its viscosity, and v the object's speed in the fluid. If $N\prime_{\mathrm{R}}$ is less than about 1, flow around the object can be laminar, particularly if the object has a smooth shape. The transition to turbulent flow occurs for $N\prime_{\mathrm{R}}$ between 1 and about 10, depending on surface roughness and so on. Depending on the surface, there can be a *turbulent wake* behind the object with some laminar flow over its surface. For an $N\prime_{\mathrm{R}}$ between 10 and 10^6, the flow may be either laminar or turbulent and may oscillate between the two. For $N\prime_{\mathrm{R}}$ greater than about 10^6, the flow is entirely turbulent, even at the surface of the object. (See [link].) Laminar flow occurs mostly when the objects in the fluid are small, such as raindrops, pollen, and blood cells in plasma.

Example:
Does a Ball Have a Turbulent Wake?

Calculate the Reynolds number $N\prime_{\mathrm{R}}$ for a ball with a 7.40-cm diameter thrown at 40.0 m/s.

Strategy

We can use $N\prime_{\mathrm{R}} = \frac{\rho \mathrm{vL}}{\eta}$ to calculate $N\prime_{\mathrm{R}}$, since all values in it are either given or can be found in tables of density and viscosity.

Solution

Substituting values into the equation for $N\prime_{\mathrm{R}}$ yields

Equation:

$$\begin{aligned} N\prime_R &= \frac{\rho \mathrm{vL}}{\eta} = \frac{(1.29\ \mathrm{kg/m^3})(40.0\ \mathrm{m/s})(0.0740\ \mathrm{m})}{1.81\times10^{-5}\ 1.00\ \mathrm{Pa\cdot s}} \\ &= 2.11 \times 10^5. \end{aligned}$$

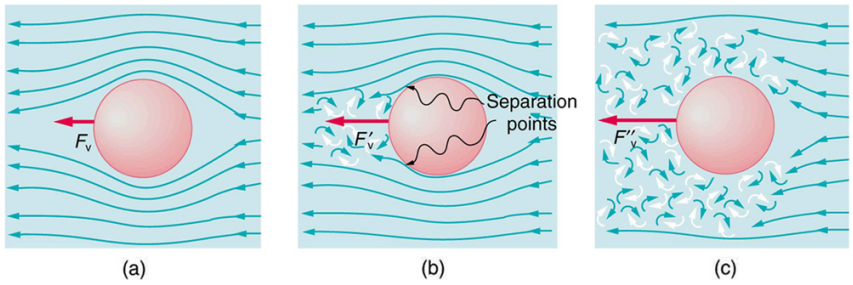
Discussion

This value is sufficiently high to imply a turbulent wake. Most large objects, such as airplanes and sailboats, create significant turbulence as they move. As noted before, the Bernoulli principle gives only qualitatively-correct results in such situations.

One of the consequences of viscosity is a resistance force called **viscous drag** F_{V} that is exerted on a moving object. This force typically depends on the object's speed (in contrast with simple friction). Experiments have shown that for laminar flow ($N\prime_{\mathrm{R}}$ less than about one) viscous drag is proportional to speed, whereas for $N\prime_{\mathrm{R}}$ between about 10 and 10^6, viscous drag is proportional to speed squared. (This relationship is a strong dependence and is pertinent to bicycle racing, where even a small headwind causes significantly increased drag on the racer. Cyclists take turns being the leader in the pack for this reason.) For $N\prime_{\mathrm{R}}$ greater than 10^6, drag increases dramatically and behaves with greater complexity. For laminar flow around a sphere, F_{V} is proportional to fluid viscosity η, the object's characteristic size L, and its speed v. All of which makes sense—the more viscous the fluid and the larger the object, the more drag we expect. Recall Stoke's law $F_{\mathrm{S}} = 6\pi r\eta v$. For the special case of a small sphere of radius R moving slowly in a fluid of viscosity η, the drag force F_{S} is given by

Equation:

$$F_{\mathrm{S}} = 6\pi R\eta v.$$



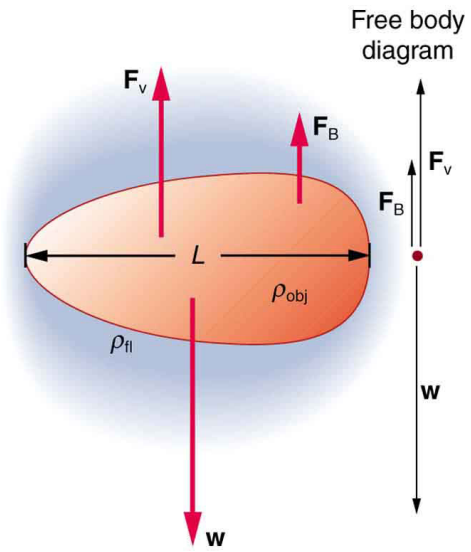
(a) Motion of this sphere to the right is equivalent to fluid flow to the left. Here the flow is laminar with $N\prime_{\mathrm{R}}$ less than 1. There is a force, called viscous drag F_{V}, to the left on the ball due to the fluid's viscosity. (b) At a higher speed, the flow becomes partially turbulent, creating a wake starting where the flow lines separate from the surface. Pressure in the wake is less than in front of the sphere, because fluid speed is less, creating a net force to the left $F\prime_{\mathrm{V}}$ that is significantly greater than for laminar flow. Here $N\prime_{\mathrm{R}}$ is greater than 10. (c) At much higher speeds, where $N\prime_{\mathrm{R}}$ is greater than 10^6, flow becomes turbulent everywhere on the surface and behind the sphere. Drag increases dramatically.

An interesting consequence of the increase in F_{V} with speed is that an object falling through a fluid will not continue to accelerate indefinitely (as it would if we neglect air resistance, for example). Instead, viscous drag increases, slowing acceleration, until a critical speed, called the **terminal speed**, is reached and the acceleration of the object becomes zero. Once this happens, the object continues to fall at constant speed (the terminal speed). This is the case for particles of sand falling in the ocean, cells falling in a centrifuge, and sky divers falling through the air. [link] shows some of the factors that affect terminal speed. There is a viscous drag on the object that

depends on the viscosity of the fluid and the size of the object. But there is also a buoyant force that depends on the density of the object relative to the fluid. Terminal speed will be greatest for low-viscosity fluids and objects with high densities and small sizes. Thus a skydiver falls more slowly with outspread limbs than when they are in a pike position—head first with hands at their side and legs together.

Note:
Take-Home Experiment: Don’t Lose Your Marbles
By measuring the terminal speed of a slowly moving sphere in a viscous fluid, one can find the viscosity of that fluid (at that temperature). It can be difficult to find small ball bearings around the house, but a small marble will do. Gather two or three fluids (syrup, motor oil, honey, olive oil, etc.) and a thick, tall clear glass or vase. Drop the marble into the center of the fluid and time its fall (after letting it drop a little to reach its terminal speed). Compare your values for the terminal speed and see if they are inversely proportional to the viscosities as listed in [link]. Does it make a difference if the marble is dropped near the side of the glass?

Knowledge of terminal speed is useful for estimating sedimentation rates of small particles. We know from watching mud settle out of dirty water that sedimentation is usually a slow process. Centrifuges are used to speed sedimentation by creating accelerated frames in which gravitational acceleration is replaced by centripetal acceleration, which can be much greater, increasing the terminal speed.



There are three forces acting on an object falling through a viscous fluid: its weight w, the viscous drag F_{V}, and the buoyant force $\mathbf{F}_{\mathrm{B}}$.

Section Summary

- When an object moves in a fluid, there is a different form of the Reynolds number $N\prime_{\mathrm{R}} = \frac{\rho \mathrm{vL}}{\eta}\,(\text{object in fluid})$, which indicates whether flow is laminar or turbulent.
- For $N\prime_{\mathrm{R}}$ less than about one, flow is laminar.
- For $N\prime_{\mathrm{R}}$ greater than 10^6, flow is entirely turbulent.

Conceptual Questions

Exercise:

Problem:

What direction will a helium balloon move inside a car that is slowing down—toward the front or back? Explain your answer.

Exercise:

Problem:

Will identical raindrops fall more rapidly in $5^{\circ}\ \mathrm{C}$ air or $25^{\circ}\ \mathrm{C}$ air, neglecting any differences in air density? Explain your answer.

Exercise:

Problem:

If you took two marbles of different sizes, what would you expect to observe about the relative magnitudes of their terminal velocities?

Glossary

viscous drag
: a resistance force exerted on a moving object, with a nontrivial dependence on velocity

terminal speed
: the speed at which the viscous drag of an object falling in a viscous fluid is equal to the other forces acting on the object (such as gravity), so that the acceleration of the object is zero

Molecular Transport Phenomena: Diffusion, Osmosis, and Related Processes

- Define diffusion, osmosis, dialysis, and active transport.
- Calculate diffusion rates.

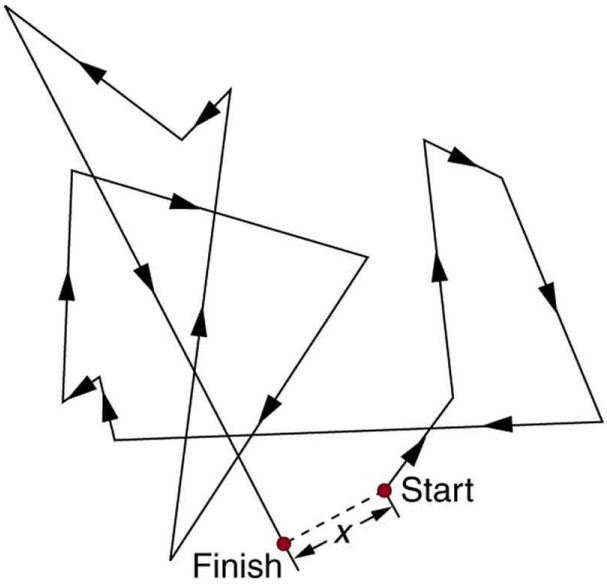
Diffusion

There is something fishy about the ice cube from your freezer—how did it pick up those food odors? How does soaking a sprained ankle in Epsom salt reduce swelling? The answer to these questions are related to atomic and molecular transport phenomena—another mode of fluid motion. Atoms and molecules are in constant motion at any temperature. In fluids they move about randomly even in the absence of macroscopic flow. This motion is called a random walk and is illustrated in [link]. **Diffusion** is the movement of substances due to random thermal molecular motion. Fluids, like fish fumes or odors entering ice cubes, can even diffuse through solids.

Diffusion is a slow process over macroscopic distances. The densities of common materials are great enough that molecules cannot travel very far before having a collision that can scatter them in any direction, including straight backward. It can be shown that the average distance x_{rms} that a molecule travels is proportional to the square root of time:
Equation:

$$x_{\text{rms}} = \sqrt{2\text{Dt}},$$

where x_{rms} stands for the **root-mean-square distance** and is the statistical average for the process. The quantity D is the diffusion constant for the particular molecule in a specific medium. [link] lists representative values of D for various substances, in units of m^2/s.



The random thermal motion of a molecule in a fluid in time t. This type of motion is called a random walk.

Diffusing molecule	**Medium**	***D*** **(m^2/s)**
Hydrogen (H_2)	Air	6.4×10^{-5}
Oxygen (O_2)	Air	1.8×10^{-5}
Oxygen (O_2)	Water	1.0×10^{-9}
Glucose $(C_6H_{12}O_6)$	Water	6.7×10^{-10}
Hemoglobin	Water	6.9×10^{-11}

Diffusing molecule	Medium	D (m^2/s)
DNA	Water	1.3×10^{-12}

Diffusion Constants for Various Molecules[footnote]
At 20°C and 1 atm

Note that D gets progressively smaller for more massive molecules. This decrease is because the average molecular speed at a given temperature is inversely proportional to molecular mass. Thus the more massive molecules diffuse more slowly. Another interesting point is that D for oxygen in air is much greater than D for oxygen in water. In water, an oxygen molecule makes many more collisions in its random walk and is slowed considerably. In water, an oxygen molecule moves only about 40 μm in 1 s. (Each molecule actually collides about 10^{10} times per second!). Finally, note that diffusion constants increase with temperature, because average molecular speed increases with temperature. This is because the average kinetic energy of molecules, $\frac{1}{2}mv^2$, is proportional to absolute temperature.

Example:
Calculating Diffusion: How Long Does Glucose Diffusion Take?
Calculate the average time it takes a glucose molecule to move 1.0 cm in water.
Strategy
We can use $x_{\text{rms}} = \sqrt{2D\text{t}}$, the expression for the average distance moved in time t, and solve it for t. All other quantities are known.
Solution
Solving for t and substituting known values yields
Equation:

$$\begin{aligned} t &= \frac{x_{\text{rms}}^2}{2D} = \frac{(0.010\ \text{m})^2}{2(6.7\times10^{-10}\ \text{m}^2/\text{s})} \\ &= 7.5 \times 10^4\ \text{s} = 21\ \text{h}. \end{aligned}$$

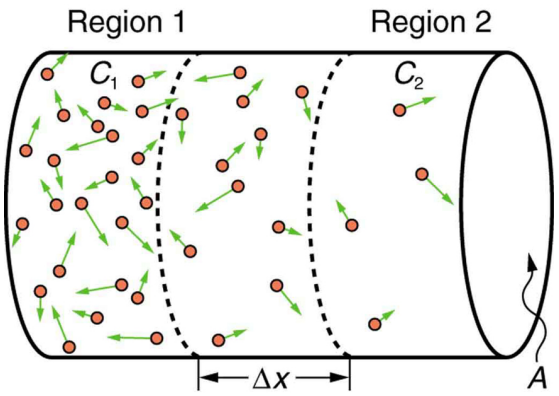
Discussion

This is a remarkably long time for glucose to move a mere centimeter! For this reason, we stir sugar into water rather than waiting for it to diffuse.

Because diffusion is typically very slow, its most important effects occur over small distances. For example, the cornea of the eye gets most of its oxygen by diffusion through the thin tear layer covering it.

The Rate and Direction of Diffusion

If you very carefully place a drop of food coloring in a still glass of water, it will slowly diffuse into the colorless surroundings until its concentration is the same everywhere. This type of diffusion is called free diffusion, because there are no barriers inhibiting it. Let us examine its direction and rate. Molecular motion is random in direction, and so simple chance dictates that more molecules will move out of a region of high concentration than into it. The net rate of diffusion is higher initially than after the process is partially completed. (See [link].)



Diffusion proceeds from a region of higher concentration to a lower one. The net rate of movement is proportional to the difference in concentration.

The net rate of diffusion is proportional to the concentration difference. Many more molecules will leave a region of high concentration than will enter it from a region of low concentration. In fact, if the concentrations were the same, there would be *no* net movement. The net rate of diffusion is also proportional to the diffusion constant D, which is determined experimentally. The farther a molecule can diffuse in a given time, the more likely it is to leave the region of high concentration. Many of the factors that affect the rate are hidden in the diffusion constant D. For example, temperature and cohesive and adhesive forces all affect values of D.

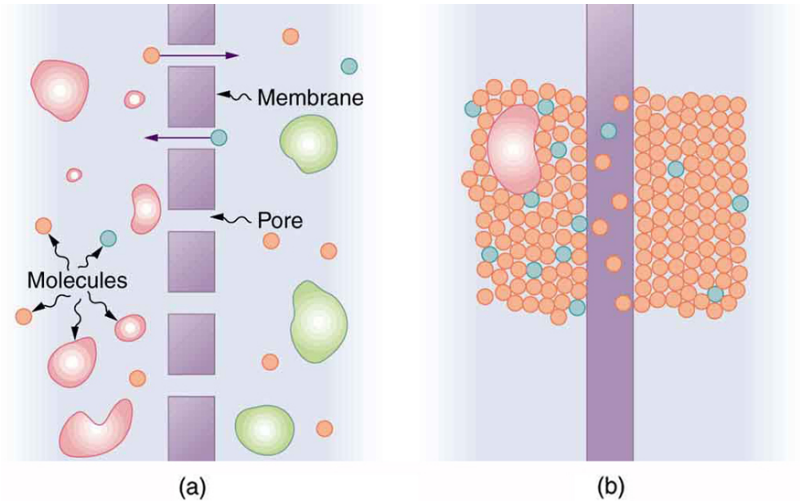
Diffusion is the dominant mechanism by which the exchange of nutrients and waste products occur between the blood and tissue, and between air and blood in the lungs. In the evolutionary process, as organisms became larger, they needed quicker methods of transportation than net diffusion, because of the larger distances involved in the transport, leading to the development of circulatory systems. Less sophisticated, single-celled organisms still rely totally on diffusion for the removal of waste products and the uptake of nutrients.

Osmosis and Dialysis—Diffusion across Membranes

Some of the most interesting examples of diffusion occur through barriers that affect the rates of diffusion. For example, when you soak a swollen ankle in Epsom salt, water diffuses through your skin. Many substances regularly move through cell membranes; oxygen moves in, carbon dioxide moves out, nutrients go in, and wastes go out, for example. Because membranes are thin structures (typically 6.5×10^{-9} to 10×10^{-9} m across) diffusion rates through them can be high. Diffusion through membranes is an important method of transport.

Membranes are generally selectively permeable, or **semipermeable**. (See [link].) One type of semipermeable membrane has small pores that allow only small molecules to pass through. In other types of membranes, the molecules may actually dissolve in the membrane or react with molecules

in the membrane while moving across. Membrane function, in fact, is the subject of much current research, involving not only physiology but also chemistry and physics.

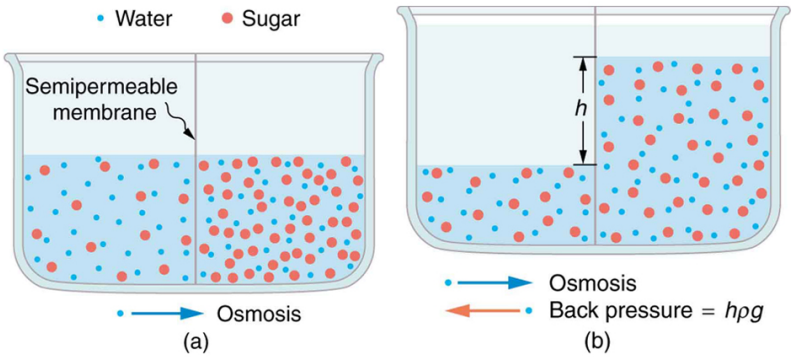


(a) A semipermeable membrane with small pores that allow only small molecules to pass through. (b) Certain molecules dissolve in this membrane and diffuse across it.

Osmosis is the transport of water through a semipermeable membrane from a region of high concentration to a region of low concentration. Osmosis is driven by the imbalance in water concentration. For example, water is more concentrated in your body than in Epsom salt. When you soak a swollen ankle in Epsom salt, the water moves out of your body into the lower-concentration region in the salt. Similarly, **dialysis** is the transport of any other molecule through a semipermeable membrane due to its concentration difference. Both osmosis and dialysis are used by the kidneys to cleanse the blood.

Osmosis can create a substantial pressure. Consider what happens if osmosis continues for some time, as illustrated in [link]. Water moves by osmosis from the left into the region on the right, where it is less

concentrated, causing the solution on the right to rise. This movement will continue until the pressure ρgh created by the extra height of fluid on the right is large enough to stop further osmosis. This pressure is called a *back pressure*. The back pressure ρgh that stops osmosis is also called the **relative osmotic pressure** if neither solution is pure water, and it is called the **osmotic pressure** if one solution is pure water. Osmotic pressure can be large, depending on the size of the concentration difference. For example, if pure water and sea water are separated by a semipermeable membrane that passes no salt, osmotic pressure will be 25.9 atm. This value means that water will diffuse through the membrane until the salt water surface rises 268 m above the pure-water surface! One example of pressure created by osmosis is turgor in plants (many wilt when too dry). Turgor describes the condition of a plant in which the fluid in a cell exerts a pressure against the cell wall. This pressure gives the plant support. Dialysis can similarly cause substantial pressures.



(a) Two sugar-water solutions of different concentrations, separated by a semipermeable membrane that passes water but not sugar. Osmosis will be to the right, since water is less concentrated there. (b) The fluid level rises until the back pressure ρgh equals the relative osmotic pressure; then, the net transfer of water is zero.

Reverse osmosis and **reverse dialysis** (also called filtration) are processes that occur when back pressure is sufficient to reverse the normal direction of substances through membranes. Back pressure can be created naturally as on the right side of [link]. (A piston can also create this pressure.) Reverse osmosis can be used to desalinate water by simply forcing it through a membrane that will not pass salt. Similarly, reverse dialysis can be used to filter out any substance that a given membrane will not pass.

One further example of the movement of substances through membranes deserves mention. We sometimes find that substances pass in the direction opposite to what we expect. Cypress tree roots, for example, extract pure water from salt water, although osmosis would move it in the opposite direction. This is not reverse osmosis, because there is no back pressure to cause it. What is happening is called **active transport**, a process in which a living membrane expends energy to move substances across it. Many living membranes move water and other substances by active transport. The kidneys, for example, not only use osmosis and dialysis—they also employ significant active transport to move substances into and out of blood. In fact, it is estimated that at least 25% of the body's energy is expended on active transport of substances at the cellular level. The study of active transport carries us into the realms of microbiology, biophysics, and biochemistry and it is a fascinating application of the laws of nature to living structures.

Section Summary

- Diffusion is the movement of substances due to random thermal molecular motion.
- The average distance x_{rms} a molecule travels by diffusion in a given amount of time is given by
  Equation:

$$x_{\text{rms}} = \sqrt{2D\text{t}},$$

  where D is the diffusion constant, representative values of which are found in [link].

- Osmosis is the transport of water through a semipermeable membrane from a region of high concentration to a region of low concentration.
- Dialysis is the transport of any other molecule through a semipermeable membrane due to its concentration difference.
- Both processes can be reversed by back pressure.
- Active transport is a process in which a living membrane expends energy to move substances across it.

Conceptual Questions

Exercise:

Problem:

Why would you expect the rate of diffusion to increase with temperature? Can you give an example, such as the fact that you can dissolve sugar more rapidly in hot water?

Exercise:

Problem: How are osmosis and dialysis similar? How do they differ?

Problem Exercises

Exercise:

Problem:

You can smell perfume very shortly after opening the bottle. To show that it is not reaching your nose by diffusion, calculate the average distance a perfume molecule moves in one second in air, given its diffusion constant D to be $1.00 \times 10^{-6}\ \mathrm{m^2/s}$.

Solution:

$1.41 \times 10^{-3}\ \mathrm{m}$

Exercise:

Problem:

What is the ratio of the average distances that oxygen will diffuse in a given time in air and water? Why is this distance less in water (equivalently, why is D less in water)?

Exercise:

Problem:

Oxygen reaches the veinless cornea of the eye by diffusing through its tear layer, which is 0.500-mm thick. How long does it take the average oxygen molecule to do this?

Solution:

1.3×10^2 s

Exercise:

Problem:

(a) Find the average time required for an oxygen molecule to diffuse through a 0.200-mm-thick tear layer on the cornea. (b) How much time is required to diffuse $0.500\ \text{cm}^3$ of oxygen to the cornea if its surface area is $1.00\ \text{cm}^2$?

Exercise:

Problem:

Suppose hydrogen and oxygen are diffusing through air. A small amount of each is released simultaneously. How much time passes before the hydrogen is 1.00 s ahead of the oxygen? Such differences in arrival times are used as an analytical tool in gas chromatography.

Solution:

0.391 s

Glossary

diffusion
: the movement of substances due to random thermal molecular motion

semipermeable
: a type of membrane that allows only certain small molecules to pass through

osmosis
: the transport of water through a semipermeable membrane from a region of high concentration to one of low concentration

dialysis
: the transport of any molecule other than water through a semipermeable membrane from a region of high concentration to one of low concentration

relative osmotic pressure
: the back pressure which stops the osmotic process if neither solution is pure water

osmotic pressure
: the back pressure which stops the osmotic process if one solution is pure water

reverse osmosis
: the process that occurs when back pressure is sufficient to reverse the normal direction of osmosis through membranes

reverse dialysis
: the process that occurs when back pressure is sufficient to reverse the normal direction of dialysis through membranes

active transport
: the process in which a living membrane expends energy to move substances across

The Cell Membrane
By the end of this section, you will be able to:

- Describe the molecular components that make up the cell membrane
- Explain the major features and properties of the cell membrane
- Differentiate between materials that can and cannot diffuse through the lipid bilayer
- Compare and contrast different types of passive transport with active transport, providing examples of each

Despite differences in structure and function, all living cells in multicellular organisms have a surrounding cell membrane. As the outer layer of your skin separates your body from its environment, the cell membrane (also known as the plasma membrane) separates the inner contents of a cell from its exterior environment. This cell membrane provides a protective barrier around the cell and regulates which materials can pass in or out.

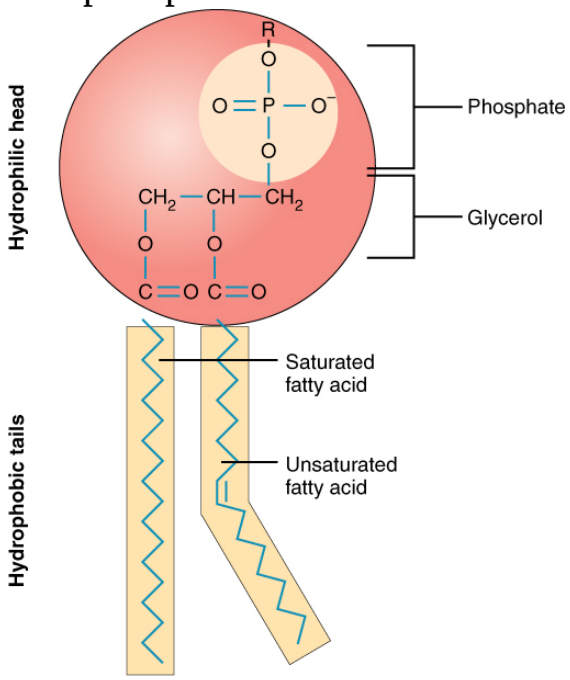
Structure and Composition of the Cell Membrane

The **cell membrane** is an extremely pliable structure composed primarily of back-to-back phospholipids (a “bilayer”). Cholesterol is also present, which contributes to the fluidity of the membrane, and there are various proteins embedded within the membrane that have a variety of functions.

A single phospholipid molecule has a phosphate group on one end, called the “head,” and two side-by-side chains of fatty acids that make up the lipid tails ([link]). The phosphate group is negatively charged, making the head polar and hydrophilic—or “water loving.” A **hydrophilic** molecule (or region of a molecule) is one that is attracted to water. The phosphate heads are thus attracted to the water molecules of both the extracellular and intracellular environments. The lipid tails, on the other hand, are uncharged, or nonpolar, and are hydrophobic—or “water fearing.” A **hydrophobic** molecule (or region of a molecule) repels and is repelled by water. Some lipid tails consist of saturated fatty acids and some contain unsaturated fatty acids. This combination adds to the fluidity of the tails that are constantly in motion. Phospholipids are thus amphipathic molecules. An **amphipathic** molecule is one that contains both a hydrophilic and a hydrophobic region.

In fact, soap works to remove oil and grease stains because it has amphipathic properties. The hydrophilic portion can dissolve in water while the hydrophobic portion can trap grease in micelles that then can be washed away.

Phospholipid Structure

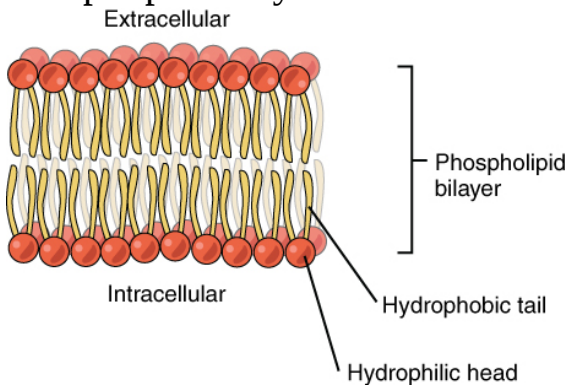


A phospholipid molecule consists of a polar phosphate "head," which is hydrophilic and a non-polar lipid "tail," which is hydrophobic. Unsaturated fatty acids result in kinks in the hydrophobic tails.

The cell membrane consists of two adjacent layers of phospholipids. The lipid tails of one layer face the lipid tails of the other layer, meeting at the interface of the two layers. The phospholipid heads face outward, one layer exposed to the interior of the cell and one layer exposed to the exterior ([link]). Because the phosphate groups are polar and hydrophilic, they are

attracted to water in the intracellular fluid. **Intracellular fluid (ICF)** is the fluid interior of the cell. The phosphate groups are also attracted to the extracellular fluid. **Extracellular fluid (ECF)** is the fluid environment outside the enclosure of the cell membrane. **Interstitial fluid (IF)** is the term given to extracellular fluid not contained within blood vessels. Because the lipid tails are hydrophobic, they meet in the inner region of the membrane, excluding watery intracellular and extracellular fluid from this space. The cell membrane has many proteins, as well as other lipids (such as cholesterol), that are associated with the phospholipid bilayer. An important feature of the membrane is that it remains fluid; the lipids and proteins in the cell membrane are not rigidly locked in place.

Phospolipid Bilayer

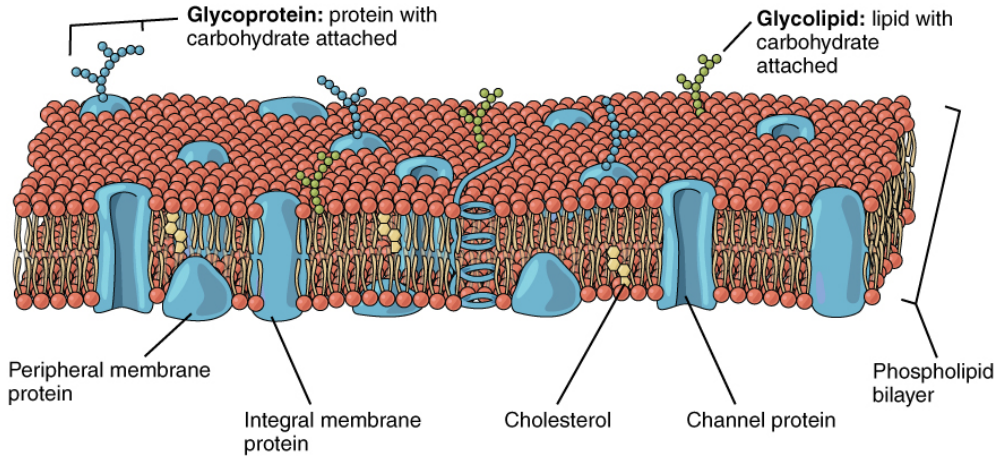


The phospholipid bilayer consists of two adjacent sheets of phospholipids, arranged tail to tail. The hydrophobic tails associate with one another, forming the interior of the membrane. The polar heads contact the fluid inside and outside of the cell.

Membrane Proteins

The lipid bilayer forms the basis of the cell membrane, but it is peppered throughout with various proteins. Two different types of proteins that are commonly associated with the cell membrane are the integral proteins and peripheral protein ([link]). As its name suggests, an **integral protein** is a protein that is embedded in the membrane. A **channel protein** is an example of an integral protein that selectively allows particular materials, such as certain ions, to pass into or out of the cell.

Cell Membrane



The cell membrane of the cell is a phospholipid bilayer containing many different molecular components, including proteins and cholesterol, some with carbohydrate groups attached.

Another important group of integral proteins are cell recognition proteins, which serve to mark a cell's identity so that it can be recognized by other cells. A **receptor** is a type of recognition protein that can selectively bind a specific molecule outside the cell, and this binding induces a chemical reaction within the cell. A **ligand** is the specific molecule that binds to and activates a receptor. Some integral proteins serve dual roles as both a receptor and an ion channel. One example of a receptor-ligand interaction is the receptors on nerve cells that bind neurotransmitters, such as dopamine. When a dopamine molecule binds to a dopamine receptor protein, a channel within the transmembrane protein opens to allow certain ions to flow into the cell.

Some integral membrane proteins are glycoproteins. A **glycoprotein** is a protein that has carbohydrate molecules attached, which extend into the extracellular matrix. The attached carbohydrate tags on glycoproteins aid in cell recognition. The carbohydrates that extend from membrane proteins and even from some membrane lipids collectively form the glycocalyx. The **glycocalyx** is a fuzzy-appearing coating around the cell formed from glycoproteins and other carbohydrates attached to the cell membrane. The glycocalyx can have various roles. For example, it may have molecules that allow the cell to bind to another cell, it may contain receptors for hormones, or it might have enzymes to break down nutrients. The glycocalyces found in a person's body are products of that person's genetic makeup. They give each of the individual's trillions of cells the "identity" of belonging in the person's body. This identity is the primary way that a person's immune defense cells "know" not to attack the person's own body cells, but it also is the reason organs donated by another person might be rejected.

Peripheral proteins are typically found on the inner or outer surface of the lipid bilayer but can also be attached to the internal or external surface of an integral protein. These proteins typically perform a specific function for the cell. Some peripheral proteins on the surface of intestinal cells, for example, act as digestive enzymes to break down nutrients to sizes that can pass through the cells and into the bloodstream.

Transport across the Cell Membrane

One of the great wonders of the cell membrane is its ability to regulate the concentration of substances inside the cell. These substances include ions such as Ca^{++}, Na^{+}, K^{+}, and Cl^{-}; nutrients including sugars, fatty acids, and amino acids; and waste products, particularly carbon dioxide (CO_2), which must leave the cell.

The membrane's lipid bilayer structure provides the first level of control. The phospholipids are tightly packed together, and the membrane has a hydrophobic interior. This structure causes the membrane to be selectively permeable. A membrane that has **selective permeability** allows only substances meeting certain criteria to pass through it unaided. In the case of the cell membrane, only relatively small, nonpolar materials can move

through the lipid bilayer (remember, the lipid tails of the membrane are nonpolar). Some examples of these are other lipids, oxygen and carbon dioxide gases, and alcohol. However, water-soluble materials—like glucose, amino acids, and electrolytes—need some assistance to cross the membrane because they are repelled by the hydrophobic tails of the phospholipid bilayer. All substances that move through the membrane do so by one of two general methods, which are categorized based on whether or not energy is required. **Passive transport** is the movement of substances across the membrane without the expenditure of cellular energy. In contrast, **active transport** is the movement of substances across the membrane using energy from adenosine triphosphate (ATP).

Passive Transport

In order to understand *how* substances move passively across a cell membrane, it is necessary to understand concentration gradients and diffusion. A **concentration gradient** is the difference in concentration of a substance across a space. Molecules (or ions) will spread/diffuse from where they are more concentrated to where they are less concentrated until they are equally distributed in that space. (When molecules move in this way, they are said to move *down* their concentration gradient.) **Diffusion** is the movement of particles from an area of higher concentration to an area of lower concentration. A couple of common examples will help to illustrate this concept. Imagine being inside a closed bathroom. If a bottle of perfume were sprayed, the scent molecules would naturally diffuse from the spot where they left the bottle to all corners of the bathroom, and this diffusion would go on until no more concentration gradient remains. Another example is a spoonful of sugar placed in a cup of tea. Eventually the sugar will diffuse throughout the tea until no concentration gradient remains. In both cases, if the room is warmer or the tea hotter, diffusion occurs even faster as the molecules are bumping into each other and spreading out faster than at cooler temperatures. Having an internal body temperature around 98.6° F thus also aids in diffusion of particles within the body.

Note:

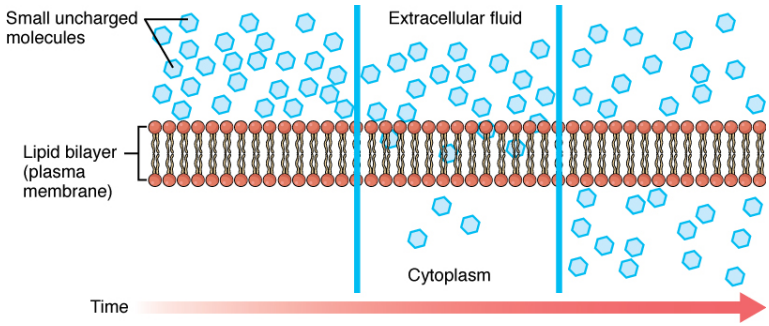


Visit this link to see diffusion and how it is propelled by the kinetic energy of molecules in solution. How does temperature affect diffusion rate, and why?

Whenever a substance exists in greater concentration on one side of a semipermeable membrane, such as the cell membranes, any substance that can move down its concentration gradient across the membrane will do so. Consider substances that can easily diffuse through the lipid bilayer of the cell membrane, such as the gases oxygen (O_2) and CO_2. O_2 generally diffuses into cells because it is more concentrated outside of them, and CO_2 typically diffuses out of cells because it is more concentrated inside of them. Neither of these examples requires any energy on the part of the cell, and therefore they use passive transport to move across the membrane.

Before moving on, you need to review the gases that can diffuse across a cell membrane. Because cells rapidly use up oxygen during metabolism, there is typically a lower concentration of O_2 inside the cell than outside. As a result, oxygen will diffuse from the interstitial fluid directly through the lipid bilayer of the membrane and into the cytoplasm within the cell. On the other hand, because cells produce CO_2 as a byproduct of metabolism, CO_2 concentrations rise within the cytoplasm; therefore, CO_2 will move from the cell through the lipid bilayer and into the interstitial fluid, where its concentration is lower. This mechanism of molecules moving across a cell membrane from the side where they are more concentrated to the side where they are less concentrated is a form of passive transport called simple diffusion ([link]).

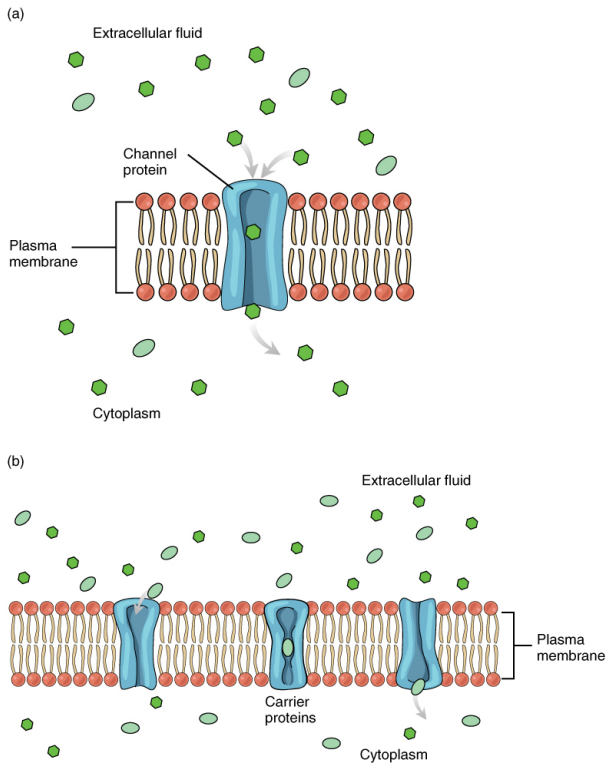
Simple Diffusion across the Cell (Plasma) Membrane



The structure of the lipid bilayer allows small, uncharged substances such as oxygen and carbon dioxide, and hydrophobic molecules such as lipids, to pass through the cell membrane, down their concentration gradient, by simple diffusion.

Large polar or ionic molecules, which are hydrophilic, cannot easily cross the phospholipid bilayer. Very small polar molecules, such as water, can cross via simple diffusion due to their small size. Charged atoms or molecules of any size cannot cross the cell membrane via simple diffusion as the charges are repelled by the hydrophobic tails in the interior of the phospholipid bilayer. Solutes dissolved in water on either side of the cell membrane will tend to diffuse down their concentration gradients, but because most substances cannot pass freely through the lipid bilayer of the cell membrane, their movement is restricted to protein channels and specialized transport mechanisms in the membrane. **Facilitated diffusion** is the diffusion process used for those substances that cannot cross the lipid bilayer due to their size, charge, and/or polarity ([link]). A common example of facilitated diffusion is the movement of glucose into the cell, where it is used to make ATP. Although glucose can be more concentrated outside of a cell, it cannot cross the lipid bilayer via simple diffusion because it is both large and polar. To resolve this, a specialized carrier protein called the glucose transporter will transfer glucose molecules into the cell to facilitate its inward diffusion.

Facilitated Diffusion



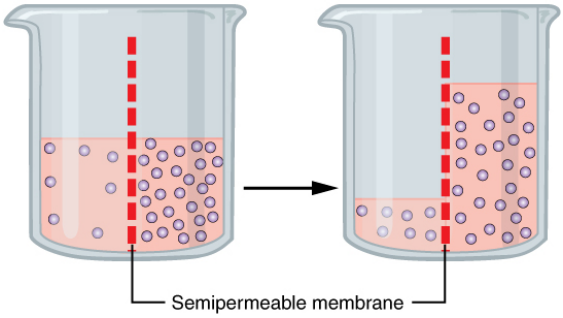
(a) Facilitated diffusion of substances crossing the cell (plasma) membrane takes place with the help of proteins such as channel proteins and carrier proteins. Channel proteins are less selective than carrier proteins, and usually mildly discriminate between their cargo based on size and charge. (b) Carrier proteins are more selective, often only allowing one particular type of molecule to cross.

As an example, even though sodium ions (Na^+) are highly concentrated outside of cells, these electrolytes are charged and cannot pass through the

nonpolar lipid bilayer of the membrane. Their diffusion is facilitated by membrane proteins that form sodium channels (or “pores”), so that Na^+ ions can move down their concentration gradient from outside the cells to inside the cells. There are many other solutes that must undergo facilitated diffusion to move into a cell, such as amino acids, or to move out of a cell, such as wastes. Because facilitated diffusion is a passive process, it does not require energy expenditure by the cell.

Water also can move freely across the cell membrane of all cells, either through protein channels or by slipping between the lipid tails of the membrane itself. **Osmosis** is the diffusion of water through a semipermeable membrane ([link]).

Osmosis

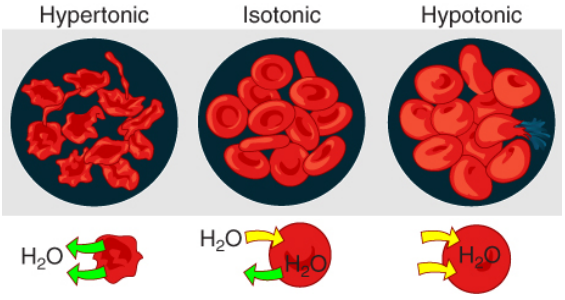


Osmosis is the diffusion of water through a semipermeable membrane down its concentration gradient. If a membrane is permeable to water, though not to a solute, water will equalize its own concentration by diffusing to the side of lower water concentration (and thus the side of higher solute concentration). In the beaker on the left, the solution on the right side of the membrane is hypertonic.

The movement of water molecules is not itself regulated by cells, so it is important that cells are exposed to an environment in which the concentration of solutes outside of the cells (in the extracellular fluid) is equal to the concentration of solutes inside the cells (in the cytoplasm). Two solutions that have the same concentration of solutes are said to be **isotonic** (equal tension). When cells and their extracellular environments are isotonic, the concentration of water molecules is the same outside and inside the cells, and the cells maintain their normal shape (and function).

Osmosis occurs when there is an imbalance of solutes outside of a cell versus inside the cell. A solution that has a higher concentration of solutes than another solution is said to be **hypertonic**, and water molecules tend to diffuse into a hypertonic solution ([link]). Cells in a hypertonic solution will shrivel as water leaves the cell via osmosis. In contrast, a solution that has a lower concentration of solutes than another solution is said to be **hypotonic**, and water molecules tend to diffuse out of a hypotonic solution. Cells in a hypotonic solution will take on too much water and swell, with the risk of eventually bursting. A critical aspect of homeostasis in living things is to create an internal environment in which all of the body's cells are in an isotonic solution. Various organ systems, particularly the kidneys, work to maintain this homeostasis.

Concentration of Solutions



A hypertonic solution has a solute concentration higher than another solution. An isotonic solution has a solute concentration equal to

another solution. A hypotonic solution has a solute concentration lower than another solution.

Another mechanism besides diffusion to passively transport materials between compartments is filtration. Unlike diffusion of a substance from where it is more concentrated to less concentrated, filtration uses a hydrostatic pressure gradient that pushes the fluid—and the solutes within it—from a higher pressure area to a lower pressure area. Filtration is an extremely important process in the body. For example, the circulatory system uses filtration to move plasma and substances across the endothelial lining of capillaries and into surrounding tissues, supplying cells with the nutrients. Filtration pressure in the kidneys provides the mechanism to remove wastes from the bloodstream.

Active Transport

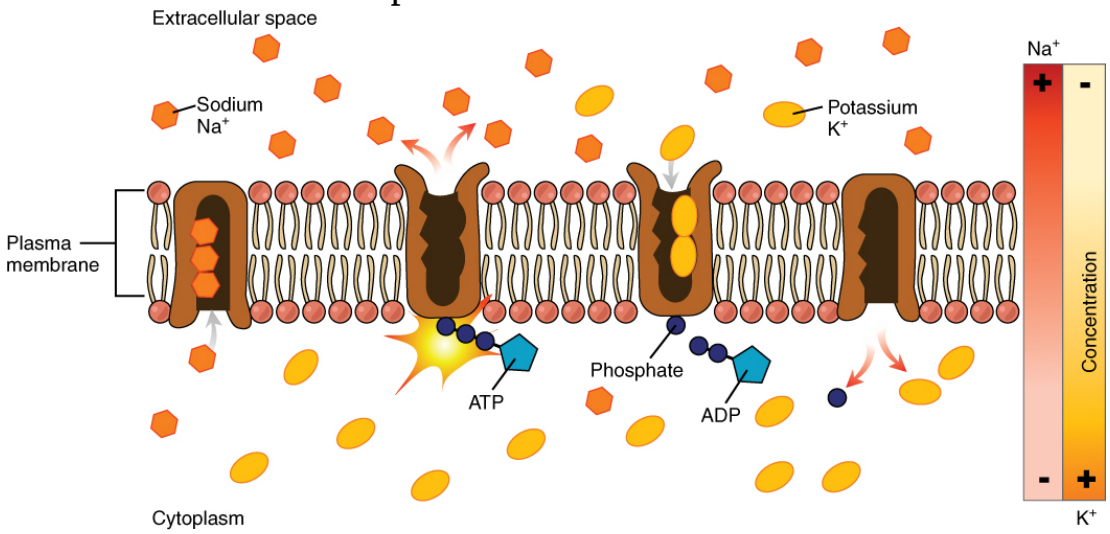
For all of the transport methods described above, the cell expends no energy. Membrane proteins that aid in the passive transport of substances do so without the use of ATP. During active transport, ATP is required to move a substance across a membrane, often with the help of protein carriers, and usually *against* its concentration gradient.

One of the most common types of active transport involves proteins that serve as pumps. The word “pump” probably conjures up thoughts of using energy to pump up the tire of a bicycle or a basketball. Similarly, energy from ATP is required for these membrane proteins to transport substances—molecules or ions—across the membrane, usually against their concentration gradients (from an area of low concentration to an area of high concentration).

The **sodium-potassium pump**, which is also called Na^+/K^+ ATPase, transports sodium out of a cell while moving potassium into the cell. The

Na^+/K^+ pump is an important ion pump found in the membranes of many types of cells. These pumps are particularly abundant in nerve cells, which are constantly pumping out sodium ions and pulling in potassium ions to maintain an electrical gradient across their cell membranes. An **electrical gradient** is a difference in electrical charge across a space. In the case of nerve cells, for example, the electrical gradient exists between the inside and outside of the cell, with the inside being negatively-charged (at around -70 mV) relative to the outside. The negative electrical gradient is maintained because each Na^+/K^+ pump moves three Na^+ ions out of the cell and two K^+ ions into the cell for each ATP molecule that is used ([link]). This process is so important for nerve cells that it accounts for the majority of their ATP usage.

Sodium-Potassium Pump



The sodium-potassium pump is found in many cell (plasma) membranes. Powered by ATP, the pump moves sodium and potassium ions in opposite directions, each against its concentration gradient. In a single cycle of the pump, three sodium ions are extruded from and two potassium ions are imported into the cell.

Active transport pumps can also work together with other active or passive transport systems to move substances across the membrane. For example, the sodium-potassium pump maintains a high concentration of sodium ions

outside of the cell. Therefore, if the cell needs sodium ions, all it has to do is open a passive sodium channel, as the concentration gradient of the sodium ions will drive them to diffuse into the cell. In this way, the action of an active transport pump (the sodium-potassium pump) powers the passive transport of sodium ions by creating a concentration gradient. When active transport powers the transport of another substance in this way, it is called secondary active transport.

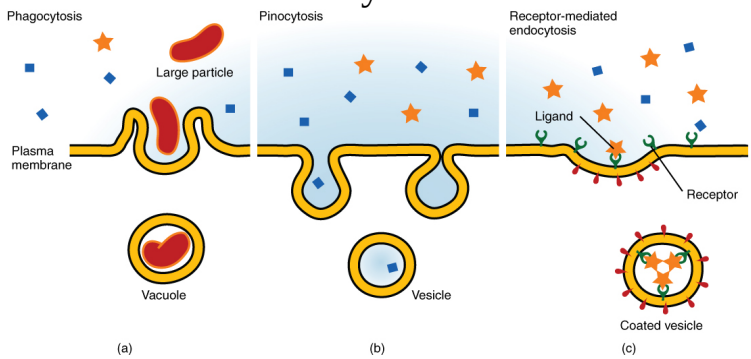
Symporters are secondary active transporters that move two substances in the same direction. For example, the sodium-glucose symporter uses sodium ions to “pull” glucose molecules into the cell. Because cells store glucose for energy, glucose is typically at a higher concentration inside of the cell than outside. However, due to the action of the sodium-potassium pump, sodium ions will easily diffuse into the cell when the symporter is opened. The flood of sodium ions through the symporter provides the energy that allows glucose to move through the symporter and into the cell, against its concentration gradient.

Conversely, antiporters are secondary active transport systems that transport substances in opposite directions. For example, the sodium-hydrogen ion antiporter uses the energy from the inward flood of sodium ions to move hydrogen ions (H^+) out of the cell. The sodium-hydrogen antiporter is used to maintain the pH of the cell's interior.

Other forms of active transport do not involve membrane carriers. **Endocytosis** (bringing “into the cell”) is the process of a cell ingesting material by enveloping it in a portion of its cell membrane, and then pinching off that portion of membrane ([link]). Once pinched off, the portion of membrane and its contents becomes an independent, intracellular vesicle. A **vesicle** is a membranous sac—a spherical and hollow organelle bounded by a lipid bilayer membrane. Endocytosis often brings materials into the cell that must to be broken down or digested. **Phagocytosis** (“cell eating”) is the endocytosis of large particles. Many immune cells engage in phagocytosis of invading pathogens. Like little Pac-men, their job is to patrol body tissues for unwanted matter, such as invading bacterial cells, phagocytize them, and digest them. In contrast to phagocytosis, **pinocytosis**

(“cell drinking”) brings fluid containing dissolved substances into a cell through membrane vesicles.

Three Forms of Endocytosis



Endocytosis is a form of active transport in which a cell envelopes extracellular materials using its cell membrane. (a) In phagocytosis, which is relatively nonselective, the cell takes in a large particle. (b) In pinocytosis, the cell takes in small particles in fluid. (c) In contrast, receptor-mediated endocytosis is quite selective. When external receptors bind a specific ligand, the cell responds by endocytosing the ligand.

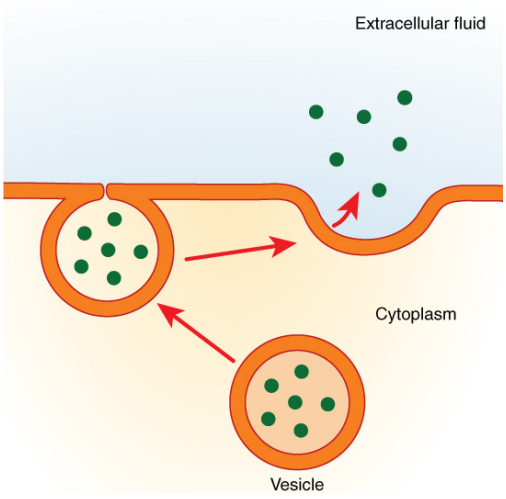
Phagocytosis and pinocytosis take in large portions of extracellular material, and they are typically not highly selective in the substances they bring in. Cells regulate the endocytosis of specific substances via receptor-mediated endocytosis. **Receptor-mediated endocytosis** is endocytosis by a portion of the cell membrane that contains many receptors that are specific for a certain substance. Once the surface receptors have bound sufficient amounts of the specific substance (the receptor’s ligand), the cell will endocytose the part of the cell membrane containing the receptor-ligand complexes. Iron, a required component of hemoglobin, is endocytosed by red blood cells in this way. Iron is bound to a protein called transferrin in the blood. Specific transferrin receptors on red blood cell surfaces bind the

iron-transferrin molecules, and the cell endocytoses the receptor-ligand complexes.

In contrast with endocytosis, **exocytosis** (taking “out of the cell”) is the process of a cell exporting material using vesicular transport ([link]). Many cells manufacture substances that must be secreted, like a factory manufacturing a product for export. These substances are typically packaged into membrane-bound vesicles within the cell. When the vesicle membrane fuses with the cell membrane, the vesicle releases it contents into the interstitial fluid. The vesicle membrane then becomes part of the cell membrane. Cells of the stomach and pancreas produce and secrete digestive enzymes through exocytosis ([link]). Endocrine cells produce and secrete hormones that are sent throughout the body, and certain immune cells produce and secrete large amounts of histamine, a chemical important for immune responses.

Exocytosis

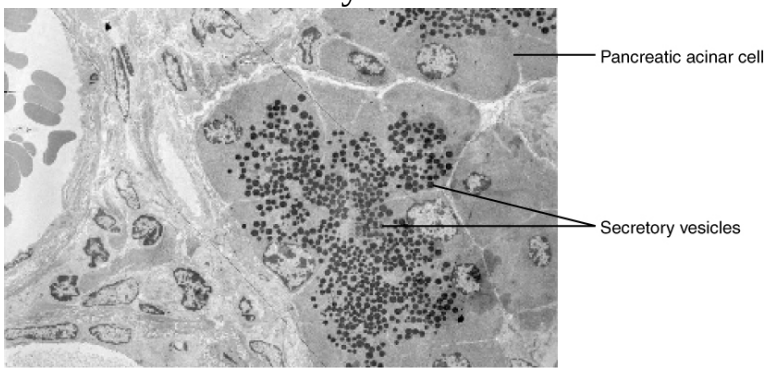
Exocytosis



Exocytosis is much like endocytosis in reverse. Material destined for export is packaged into a vesicle inside the cell. The membrane of the vesicle fuses with the cell membrane, and the

contents are released into the extracellular space.

Pancreatic Cells' Enzyme Products



The pancreatic acinar cells produce and secrete many enzymes that digest food. The tiny black granules in this electron micrograph are secretory vesicles filled with enzymes that will be exported from the cells via exocytosis. LM × 2900. (Micrograph provided by the Regents of University of Michigan Medical School © 2012)

Note:

View the University of Michigan WebScope to explore the tissue sample in greater detail.

Note:
Diseases of the…
Cell: Cystic Fibrosis
Cystic fibrosis (CF) affects approximately 30,000 people in the United States, with about 1,000 new cases reported each year. The genetic disease is most well known for its damage to the lungs, causing breathing difficulties and chronic lung infections, but it also affects the liver, pancreas, and intestines. Only about 50 years ago, the prognosis for children born with CF was very grim—a life expectancy rarely over 10 years. Today, with advances in medical treatment, many CF patients live into their 30s.
The symptoms of CF result from a malfunctioning membrane ion channel called the cystic fibrosis transmembrane conductance regulator, or CFTR. In healthy people, the CFTR protein is an integral membrane protein that transports Cl^- ions out of the cell. In a person who has CF, the gene for the CFTR is mutated, thus, the cell manufactures a defective channel protein that typically is not incorporated into the membrane, but is instead degraded by the cell.
The CFTR requires ATP in order to function, making its Cl^- transport a form of active transport. This characteristic puzzled researchers for a long time because the Cl^- ions are actually flowing *down* their concentration gradient when transported out of cells. Active transport generally pumps ions *against* their concentration gradient, but the CFTR presents an exception to this rule.
In normal lung tissue, the movement of Cl^- out of the cell maintains a Cl^--rich, negatively charged environment immediately outside of the cell. This is particularly important in the epithelial lining of the respiratory system. Respiratory epithelial cells secrete mucus, which serves to trap dust, bacteria, and other debris. A cilium (plural = cilia) is one of the hair-like appendages found on certain cells. Cilia on the epithelial cells move the mucus and its trapped particles up the airways away from the lungs and toward the outside. In order to be effectively moved upward, the mucus

cannot be too viscous; rather it must have a thin, watery consistency. The transport of Cl^- and the maintenance of an electronegative environment outside of the cell attract positive ions such as Na^+ to the extracellular space. The accumulation of both Cl^- and Na^+ ions in the extracellular space creates solute-rich mucus, which has a low concentration of water molecules. As a result, through osmosis, water moves from cells and extracellular matrix into the mucus, "thinning" it out. This is how, in a normal respiratory system, the mucus is kept sufficiently watered-down to be propelled out of the respiratory system.

If the CFTR channel is absent, Cl^- ions are not transported out of the cell in adequate numbers, thus preventing them from drawing positive ions. The absence of ions in the secreted mucus results in the lack of a normal water concentration gradient. Thus, there is no osmotic pressure pulling water into the mucus. The resulting mucus is thick and sticky, and the ciliated epithelia cannot effectively remove it from the respiratory system. Passageways in the lungs become blocked with mucus, along with the debris it carries. Bacterial infections occur more easily because bacterial cells are not effectively carried away from the lungs.

Chapter Review

The cell membrane provides a barrier around the cell, separating its internal components from the extracellular environment. It is composed of a phospholipid bilayer, with hydrophobic internal lipid "tails" and hydrophilic external phosphate "heads." Various membrane proteins are scattered throughout the bilayer, both inserted within it and attached to it peripherally. The cell membrane is selectively permeable, allowing only a limited number of materials to diffuse through its lipid bilayer. All materials that cross the membrane do so using passive (non energy-requiring) or active (energy-requiring) transport processes. During passive transport, materials move by simple diffusion or by facilitated diffusion through the membrane, down their concentration gradient. Water passes through the membrane in a diffusion process called osmosis. During active transport, energy is expended to assist material movement across the membrane in a

direction against their concentration gradient. Active transport may take place with the help of protein pumps or through the use of vesicles.

Interactive Link Questions

Exercise:

Problem:

Visit this link to see diffusion and how it is propelled by the kinetic energy of molecules in solution. How does temperature affect diffusion rate, and why?

Solution:

Higher temperatures speed up diffusion because molecules have more kinetic energy at higher temperatures.

Review Questions

Exercise:

Problem:

Because they are embedded within the membrane, ion channels are examples of _____.

a. receptor proteins
b. integral proteins
c. peripheral proteins
d. glycoproteins

Solution:

B

Exercise:

Problem:

The diffusion of substances within a solution tends to move those substances _____ their _____ gradient.

a. up; electrical
b. up; electrochemical
c. down; pressure
d. down; concentration

Solution:

D

Exercise:

Problem: Ion pumps and phagocytosis are both examples of _____.

a. endocytosis
b. passive transport
c. active transport
d. facilitated diffusion

Solution:

C

Exercise:

Problem:

Choose the answer that best completes the following analogy:
Diffusion is to _____ as endocytosis is to _____.

a. filtration; phagocytosis
b. osmosis; pinocytosis
c. solutes; fluid

d. gradient; chemical energy

Solution:

B

Critical Thinking Questions

Exercise:

Problem:

What materials can easily diffuse through the lipid bilayer, and why?

Solution:

Only materials that are relatively small and nonpolar can easily diffuse through the lipid bilayer. Large particles cannot fit in between the individual phospholipids that are packed together, and polar molecules are repelled by the hydrophobic/nonpolar lipids that line the inside of the bilayer.

Exercise:

Problem:

Why is receptor-mediated endocytosis said to be more selective than phagocytosis or pinocytosis?

Solution:

Receptor-mediated endocytosis is more selective because the substances that are brought into the cell are the specific ligands that could bind to the receptors being endocytosed. Phagocytosis or pinocytosis, on the other hand, have no such receptor-ligand specificity, and bring in whatever materials happen to be close to the membrane when it is enveloped.

Exercise:

Problem:

What do osmosis, diffusion, filtration, and the movement of ions away from like charge all have in common? In what way do they differ?

Solution:

These four phenomena are similar in the sense that they describe the movement of substances down a particular type of gradient. Osmosis and diffusion involve the movement of water and other substances down their concentration gradients, respectively. Filtration describes the movement of particles down a pressure gradient, and the movement of ions away from like charge describes their movement down their electrical gradient.

Glossary

active transport
: form of transport across the cell membrane that requires input of cellular energy

amphipathic
: describes a molecule that exhibits a difference in polarity between its two ends, resulting in a difference in water solubility

cell membrane
: membrane surrounding all animal cells, composed of a lipid bilayer interspersed with various molecules; also known as plasma membrane

channel protein
: membrane-spanning protein that has an inner pore which allows the passage of one or more substances

concentration gradient
: difference in the concentration of a substance between two regions

diffusion
: movement of a substance from an area of higher concentration to one of lower concentration

electrical gradient
: difference in the electrical charge (potential) between two regions

endocytosis
: import of material into the cell by formation of a membrane-bound vesicle

exocytosis
: export of a substance out of a cell by formation of a membrane-bound vesicle

extracellular fluid (ECF)
: fluid exterior to cells; includes the interstitial fluid, blood plasma, and fluid found in other reservoirs in the body

facilitated diffusion
: diffusion of a substance with the aid of a membrane protein

glycocalyx
: coating of sugar molecules that surrounds the cell membrane

glycoprotein
: protein that has one or more carbohydrates attached

hydrophilic
: describes a substance or structure attracted to water

hydrophobic
: describes a substance or structure repelled by water

hypertonic
: describes a solution concentration that is higher than a reference concentration

hypotonic

describes a solution concentration that is lower than a reference concentration

integral protein
: membrane-associated protein that spans the entire width of the lipid bilayer

interstitial fluid (IF)
: fluid in the small spaces between cells not contained within blood vessels

intracellular fluid (ICF)
: fluid in the cytosol of cells

isotonic
: describes a solution concentration that is the same as a reference concentration

ligand
: molecule that binds with specificity to a specific receptor molecule

osmosis
: diffusion of water molecules down their concentration gradient across a selectively permeable membrane

passive transport
: form of transport across the cell membrane that does not require input of cellular energy

peripheral protein
: membrane-associated protein that does not span the width of the lipid bilayer, but is attached peripherally to integral proteins, membrane lipids, or other components of the membrane

phagocytosis
: endocytosis of large particles

pinocytosis

endocytosis of fluid

receptor
: protein molecule that contains a binding site for another specific molecule (called a ligand)

receptor-mediated endocytosis
: endocytosis of ligands attached to membrane-bound receptors

selective permeability
: feature of any barrier that allows certain substances to cross but excludes others

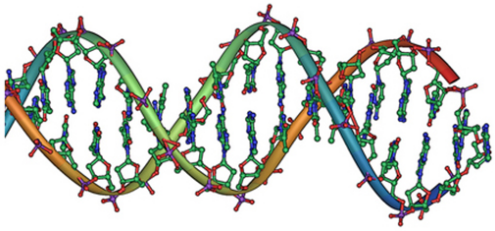
sodium-potassium pump
: (also, Na^+/K^+ ATP-ase) membrane-embedded protein pump that uses ATP to move Na^+ out of a cell and K^+ into the cell

vesicle
: membrane-bound structure that contains materials within or outside of the cell

Electric Forces in Biology

- Describe how a water molecule is polar.
- Explain electrostatic screening by a water molecule within a living cell.

Classical electrostatics has an important role to play in modern molecular biology. Large molecules such as proteins, nucleic acids, and so on—so important to life—are usually electrically charged. DNA itself is highly charged; it is the electrostatic force that not only holds the molecule together but gives the molecule structure and strength. [link] is a schematic of the DNA double helix.



DNA is a highly charged molecule. The DNA double helix shows the two coiled strands each containing a row of nitrogenous bases, which “code” the genetic information needed by a living organism. The strands are connected by bonds between pairs of bases. While pairing combinations between certain bases are fixed (C-G and A-T), the sequence of nucleotides in the strand varies. (credit: Jerome Walker)

The four nucleotide bases are given the symbols A (adenine), C (cytosine), G (guanine), and T (thymine). The order of the four bases varies in each strand, but the pairing between bases is always the same. C and G are always paired and A and T are always paired, which helps to preserve the order of bases in cell division (mitosis) so as to pass on the correct genetic information. Since the Coulomb force drops with distance ( ), the distances between the base pairs must be small enough that the electrostatic force is sufficient to hold them together.

DNA is a highly charged molecule, with about (fundamental charge) per m. The distance separating the two strands that make up the DNA structure is about 1 nm, while the distance separating the individual atoms within each base is about 0.3 nm.

One might wonder why electrostatic forces do not play a larger role in biology than they do if we have so many charged molecules. The reason is that the electrostatic force is "diluted" due to **screening** between molecules. This is due to the presence of other charges in the cell.

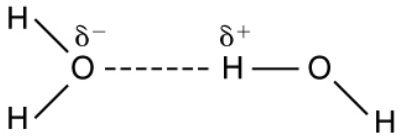
Polarity of Water Molecules

The best example of this charge screening is the water molecule, represented as . Water is a strongly **polar molecule**. Its 10 electrons (8 from the oxygen atom and 2 from the two hydrogen atoms) tend to remain closer to the oxygen nucleus than the hydrogen nuclei. This creates two centers of equal and opposite charges—what is called a **dipole**, as illustrated in [link]. The magnitude of the dipole is called the dipole moment.

These two centers of charge will terminate some of the electric field lines coming from a free charge, as on a DNA molecule. This results in a reduction in the strength of the **Coulomb interaction**. One might say that screening makes the Coulomb force a short range force rather than long range.

Other ions of importance in biology that can reduce or screen Coulomb interactions are and and . These ions are located both inside and outside of living cells. The movement of these ions through cell membranes is crucial to the motion of nerve impulses through nerve axons.

Recent studies of electrostatics in biology seem to show that electric fields in cells can be extended over larger distances, in spite of screening, by “microtubules” within the cell. These microtubules are hollow tubes composed of proteins that guide the movement of chromosomes when cells divide, the motion of other organisms within the cell, and provide mechanisms for motion of some cells (as motors).



This schematic shows water ( ) as a polar molecule. Unequal sharing of electrons between the oxygen (O) and hydrogen (H) atoms leads to a net separation of positive and negative charge—forming a dipole. The symbols and indicate that the oxygen side of the molecule tends to be more negative, while the hydrogen ends tend

to be more positive. This leads to an attraction of opposite charges between molecules.

Section Summary

- Many molecules in living organisms, such as DNA, carry a charge.
- An uneven distribution of the positive and negative charges within a polar molecule produces a dipole.
- The effect of a Coulomb field generated by a charged object may be reduced or blocked by other nearby charged objects.
- Biological systems contain water, and because water molecules are polar, they have a strong effect on other molecules in living systems.

Conceptual Question

Exercise:

Problem:

A cell membrane is a thin layer enveloping a cell. The thickness of the membrane is much less than the size of the cell. In a static situation the membrane has a charge distribution of C/m ^{2} on its inner surface and C/m^2 on its outer surface. Draw a diagram of the cell and the surrounding cell membrane. Include on this diagram the charge distribution and the corresponding electric field. Is there any electric field inside the cell? Is there any electric field outside the cell?

Glossary

dipole

a molecule’s lack of symmetrical charge distribution, causing one side to be more positive and another to be more negative

polar molecule
: a molecule with an asymmetrical distribution of positive and negative charge

screening
: the dilution or blocking of an electrostatic force on a charged object by the presence of other charges nearby

Coulomb interaction
: the interaction between two charged particles generated by the Coulomb forces they exert on one another

OU Human Physiology: The Action Potential
By the end of this section, you will be able to:

- Identify the types of gated channels
- Discuss the function of each type of gated channel
- Describe the components of the membrane that establish the resting membrane potential
- Describe the changes that occur to the membrane that result in the action potential
- Compare and contrast the configuration of the voltage gated sodium and potassium channels
- Explain why resting membrane potential is negative
- Describe how the Na^+/K^+ pump maintains the resting membrane potential and the leak channels set the resting membrane potential
- Describe each phase of the action potential including configuration of the voltage-gated sodium and potassium channels and ion permeabilities
- Compare and contrast refractory periods

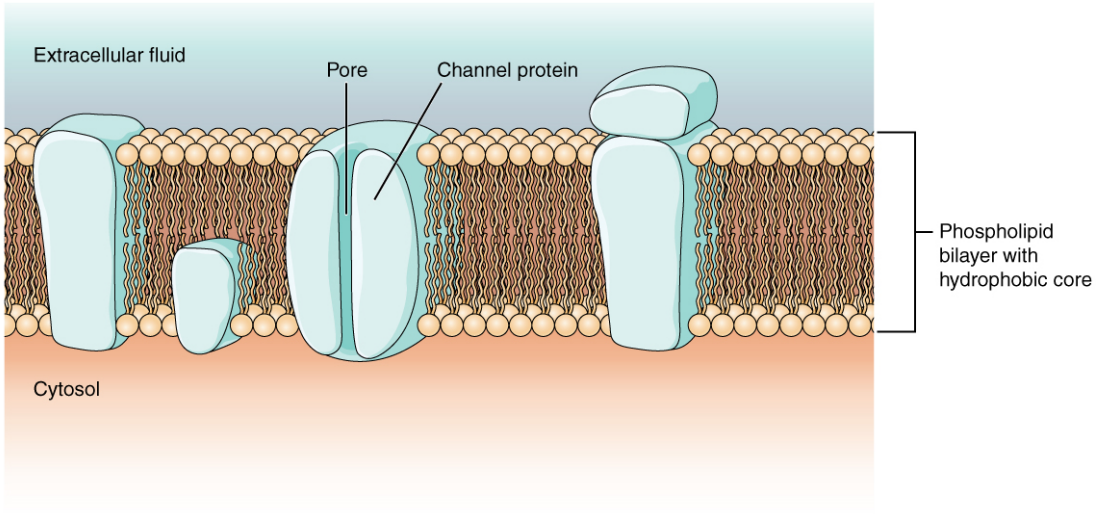
The functions of the nervous system—sensation, integration, and response—depend on the functions of the neurons underlying these pathways. To understand how neurons are able to communicate, it is necessary to describe the role of an **excitable membrane** in generating these signals. The basis of this communication is the action potential, which demonstrates how changes in the membrane can constitute a signal. Looking at the way these signals work in more variable circumstances involves a look at graded potentials, which will be covered in the next section.

Electrically Active Cell Membranes

As you learned in the chapter on cells, the cell membrane is primarily responsible for regulating what can cross the membrane and what stays on only one side. The cell membrane is a phospholipid bilayer, so only substances that can pass directly through the hydrophobic core can diffuse through unaided. Charged particles, which are hydrophilic by definition, cannot pass through the cell membrane without assistance ([link]). Transmembrane proteins, specifically channel proteins, make this possible.

Several channels, as well as specialized energy dependent “ion-pumps,” are necessary to generate a transmembrane potential and to generate an action potential. Of special interest is the carrier protein referred to as the sodium/potassium pump that moves sodium ions (Na^+) out of a cell and potassium ions (K^+) into a cell, thus maintaining ion concentration on both sides of the cell membrane.

Cell Membrane and Transmembrane Proteins



The cell membrane is composed of a phospholipid bilayer and has many transmembrane proteins, including different types of channel proteins that serve as ion channels.

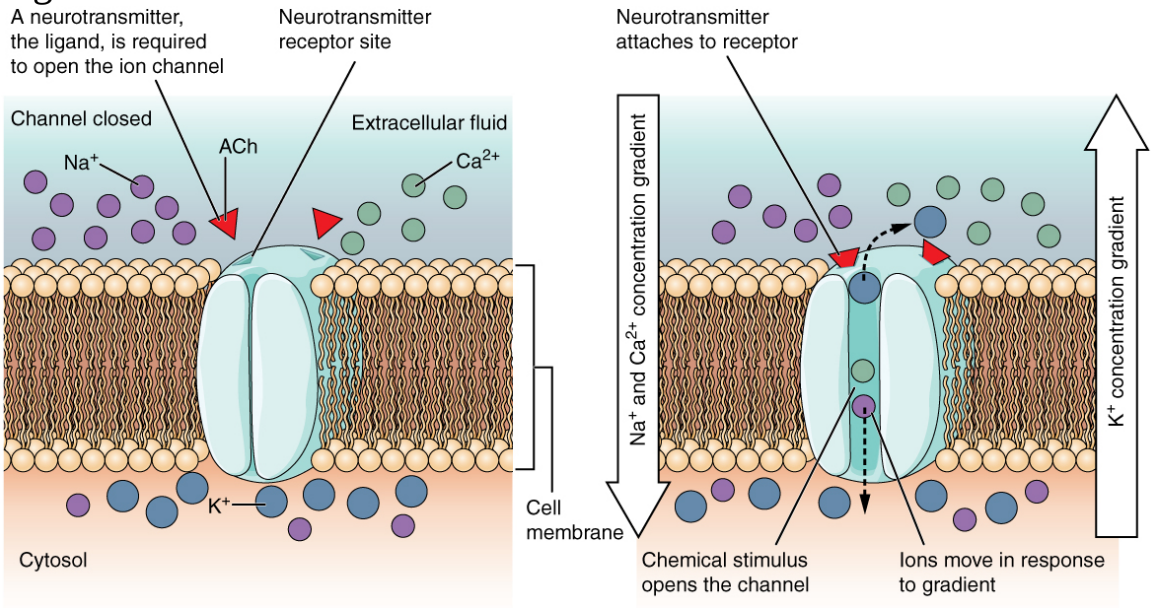
The sodium/potassium pump requires energy in the form of adenosine triphosphate (ATP), so it is also referred to as an ATPase. As was explained in the cell chapter, the concentration of Na^+ is higher outside the cell than inside, and the concentration of K^+ is higher inside the cell than outside. That means that this pump is moving the ions against the concentration gradients for sodium and potassium, which is why it requires energy. In fact, the pump basically maintains those concentration gradients.

Ion channels are pores that allow specific charged particles to cross the membrane in response to an existing gradient. Remember channels are a transmembrane protein that exhibit specificity.

Ion channels do not always freely allow ions to diffuse across the membrane. They are opened by certain events, meaning the channels are **gated**. Channels can be categorized is on the basis of how they are gated. Although these classes of ion channels are found primarily in cells of nervous or muscular tissue, they also can be found in cells of epithelial and connective tissues.

A **ligand-gated channel** opens because a signaling molecule, a ligand, binds to the extracellular region of the channel ([link]).

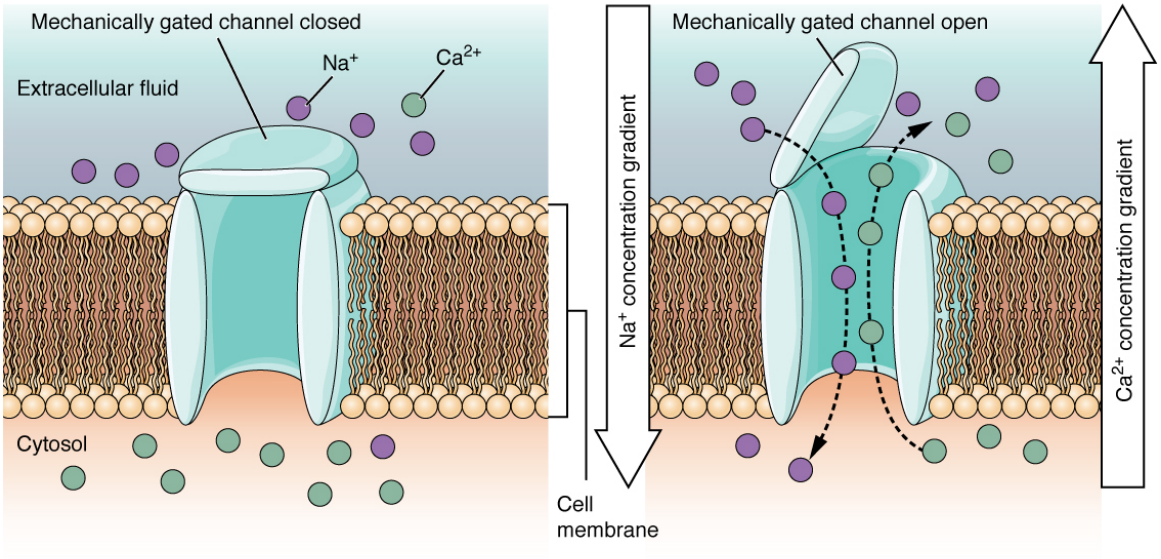
Ligand-Gated Channels



When the ligand, in this case the neurotransmitter acetylcholine, binds to a specific location on the extracellular surface of the channel protein, the pore opens to allow select ions through. The ions, in this case, are cations of sodium, calcium, and potassium.

A **mechanically gated channel** opens because of a physical distortion of the cell membrane. Many channels associated with the sense of touch (somatosensation) are mechanically gated. For example, as pressure is applied to the skin, these channels open and allow ions to enter the cell. Similar to this type of channel would be the channel that opens on the basis of temperature changes, as in testing the water in the shower ([link]).

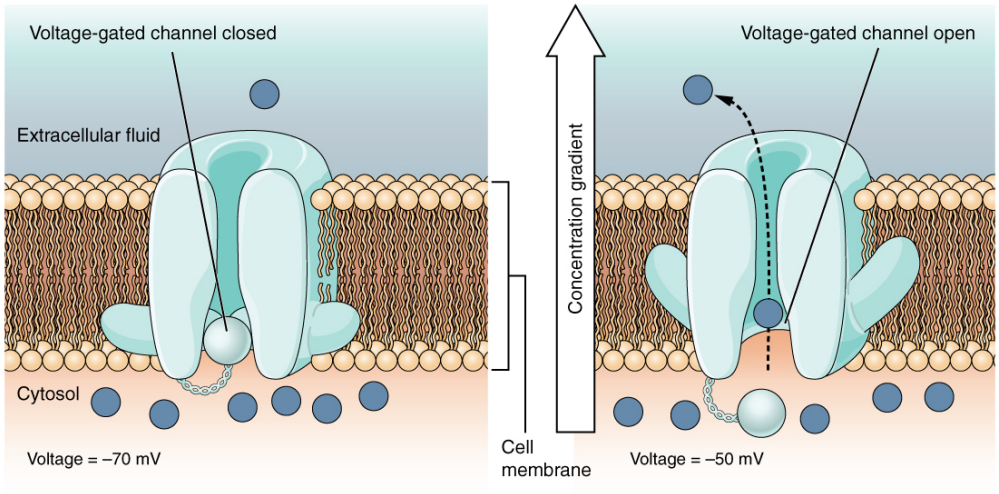
Mechanically Gated Channels



When a mechanical change occurs in the surrounding tissue, such as pressure or touch, the channel is physically opened. Thermoreceptors work on a similar principle. When the local tissue temperature changes, the protein reacts by physically opening the channel.

A **voltage-gated channel** is a channel that responds to changes in the electrical properties of the membrane in which it is embedded. Normally, the inner portion of the membrane is at a negative voltage. When that voltage becomes less negative, the channel begins to allow ions to cross the membrane ([link]).

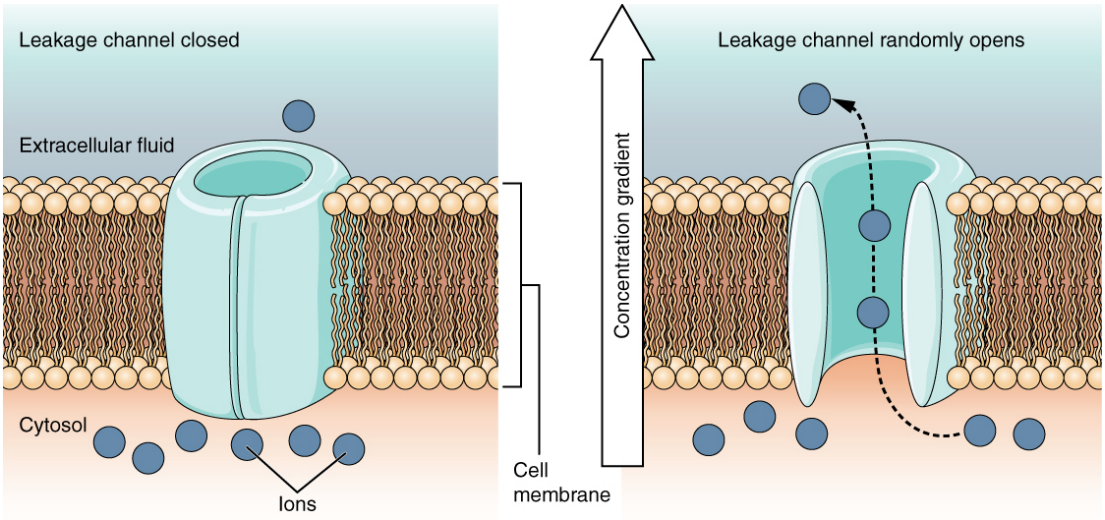
Voltage-Gated Channels



Voltage-gated channels open when the transmembrane voltage changes around them.

A **leak channel** is randomly gated, meaning that it opens and closes at random, hence the reference to leaking. There is no actual event that opens the channel; instead, it has an intrinsic rate of switching between the open and closed states. Leakage channels contribute to the resting transmembrane voltage of the excitable membrane ([link]). For our purposes, we will assume that leak channels are always open unless otherwise discussed.

Leak Channels

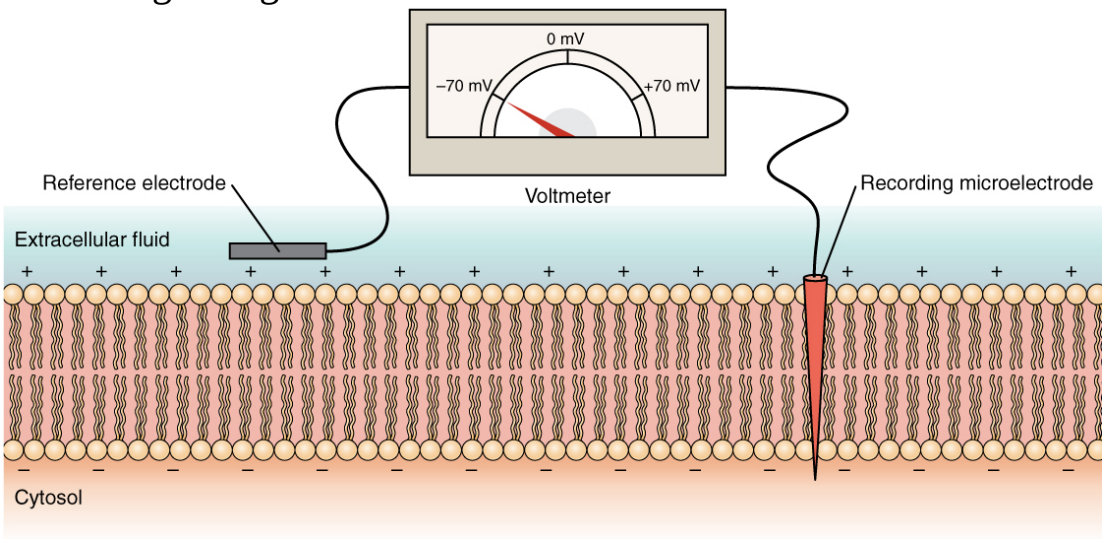


In certain situations, ions need to move across the membrane randomly. The particular electrical properties of certain cells are modified by the presence of this type of channel.

The Membrane Potential

The electrical state of the cell membrane can have several variations. These are all variations in the **membrane potential**. A potential is a distribution of charge across the cell membrane, measured in millivolts (mV). The standard is to compare the inside of the cell relative to the outside, so the membrane potential is a value representing the charge on the intracellular side of the membrane based on the outside being zero, relatively speaking ([link]).

Measuring Charge across a Membrane with a Voltmeter



A recording electrode is inserted into the cell and a reference electrode is outside the cell. By comparing the charge measured by these two electrodes, the transmembrane voltage is determined. It is conventional to express that value for the cytosol relative to the outside.

The concentration of ions in extracellular and intracellular fluids is largely balanced, with a net neutral charge. However, a slight difference in charge occurs right at the membrane surface, both internally and externally. It is the difference in this very limited region that has all the power in neurons (and muscle cells) to generate electrical signals, including action potentials.

Before these electrical signals can be described, the resting state of the membrane must be explained. When the cell is at rest (in other words it is not responding to a stimulus), and the ion channels are closed (except for leak channels), ions are distributed across the membrane in a very predictable way. The concentration of Na^+ outside the cell is 10 times greater than the concentration inside. Also, the concentration of K^+ inside the cell is greater than outside. The cytosol contains a high concentration of anions, in the form of phosphate ions and negatively charged proteins that are impermeable. Large anions are a component of the inner cell membrane, including specialized phospholipids and proteins associated with the inner leaflet of the membrane (leaflet is a term used for one side of the lipid bilayer membrane). The negative charge is localized in the large anions.

With the ions distributed across the membrane at these concentrations, the difference in charge is measured at -70 mV, the value described as the **resting membrane potential**. The exact value measured for the resting membrane potential varies between cells, but -70 mV is most commonly used as this value. This voltage would actually be much lower except for the contributions of some important proteins in the membrane. Leak channels allow Na^+ to slowly move into the cell or K^+ to slowly move out, and the Na^+/K^+ pump restores them. This may appear to be a waste of energy, but each has a role in maintaining the membrane potential.

The Action Potential

Resting membrane potential describes the steady state of the cell, which is a dynamic process that is balanced by ion leakage and ion pumping. Without any outside influence, it will not change. To get an electrical signal started, the membrane potential has to change.

This starts with a channel opening for Na^+ in the membrane. Because the concentration of Na^+ is higher outside the cell than inside the cell by a factor of 10, ions will rush into the cell that are driven by a net electrochemical gradient. Because sodium is a positively charged ion, it will change the relative voltage immediately inside the cell relative to immediately outside. The resting potential is the state of the membrane at a voltage of -70 mV, so the sodium cation entering the cell will cause it to become less negative. This is known as **depolarization**, meaning the membrane potential moves toward zero.

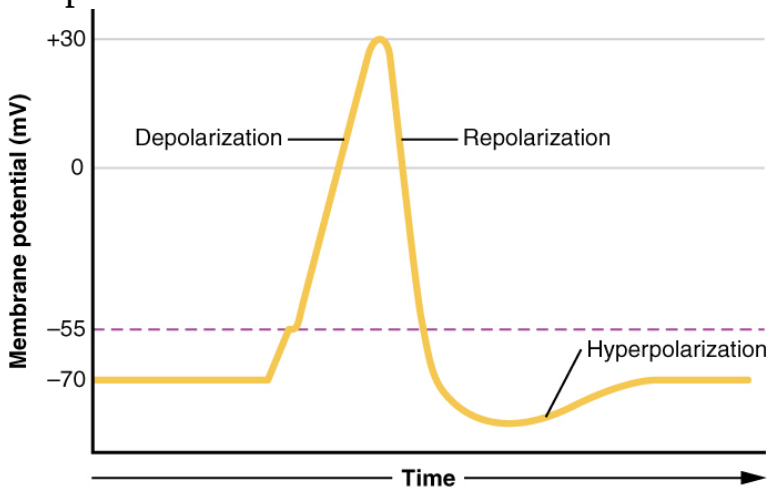
The concentration gradient for Na^+ is so strong that it will continue to enter the cell even after the membrane potential has become zero, so that the voltage immediately around the channel begins to become positive and reaches +30 mV. The membrane potential will then become more negative toward the resting voltage during repolarization. Repolarization is due to the efflux of K^+ ions along an electrochemical gradient via voltage-gated K^+ channels.

Repolarization returns the membrane potential to the -70 mV value that indicates the resting potential, but it actually overshoots that value. Potassium ions nearly reach equilibrium when the membrane voltage is below -70 mV, so a period of hyperpolarization occurs while the K^+ channels are open. Those K^+ channels are slightly delayed in closing, accounting for this short overshoot.

What has been described here is the action potential, which is presented as a graph of voltage over time in [link]. It is the electrical signal that nervous tissue generates for communication. The change in the membrane voltage from -70 mV at rest to +30 mV at the end of depolarization is a 100-mV change. That can also be written as a 0.1-V change. To put that value in perspective, think about a battery. An AA battery that you might find in a television remote has a voltage of 1.5 V, or a 9-V battery (the rectangular battery with two posts on one end) is, obviously, 9 V. The change seen in the action potential is one or two orders of magnitude less than the charge in these batteries. In fact, the membrane potential can be described as a battery. A charge is stored across the membrane that can be released under

the correct conditions. A battery in your remote has stored a charge that is “released” when you push a button.

Graph of Action Potential



Plotting voltage measured across the cell membrane against time, the action potential begins with depolarization, followed by repolarization, which goes past the resting potential into hyperpolarization, and finally the membrane returns to rest.

Note:

What happens across the membrane of an electrically active cell is a dynamic process that is hard to visualize with static images or through text descriptions. View this animation to learn more about this process. What is

the difference between the driving force for Na^+ and K^+? And what is similar about the movement of these two ions?

The question is, now, what initiates the action potential? The description above conveniently glosses over that point. But it is vital to understanding what is happening. The membrane potential will stay at the resting voltage until something changes. The description above just says that a Na^+ channel opens. Now, to say “a channel opens” does not mean that one individual transmembrane protein changes. Instead, it means that one kind of channel opens. There are a few different types of channels that allow Na^+ to cross the membrane. A ligand-gated Na^+ channel will open when a neurotransmitter binds to it and a mechanically gated Na^+ channel will open when a physical stimulus affects a sensory receptor (like pressure applied to the skin compresses a touch receptor). Whether it is a neurotransmitter binding to its receptor protein or a sensory stimulus activating a sensory receptor cell, some stimulus gets the process started. Sodium starts to enter the cell and the membrane becomes less negative.

The ligand-gated Na^+ channels that start depolarizing the membrane because of a stimulus help the cell to depolarize from -70 mV to -55 mV. Negative 55 mV is known as the threshold. Once the membrane reaches that voltage an action potential will occur. Any depolarization that does not change the membrane potential to -55 mV or higher will not reach threshold and thus will not result in an action potential.

Because of the threshold, the action potential can be likened to a digital event—it either happens or it does not. If the threshold is not reached, then no action potential occurs. If depolarization reaches -55 mV, then the action potential continues and runs all the way to +30 mV, at which K^+ causes repolarization, including the hyperpolarizing overshoot. Also, those changes are the same for every action potential, which means that once the threshold is reached, the exact same thing happens. A stronger stimulus, which might depolarize the membrane well past threshold, will not make a “bigger” action potential. Action potentials are “all or none.” Either the membrane reaches the threshold and everything occurs as described above, or the membrane does not reach the threshold and nothing else happens. All

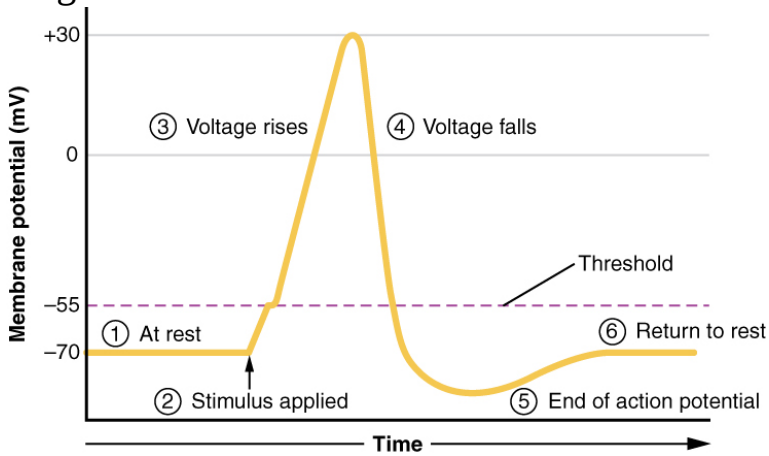
action potentials peak at the same voltage (+30 mV), so one action potential is not bigger than another. Stronger stimuli will initiate multiple action potentials more quickly, but the individual signals are not bigger. Thus, for example, you will not feel a greater sensation of pain, or have a stronger muscle contraction, because of the size of the action potential because they are not different sizes.

The depolarization and repolarization of an action potential are dependent on two types of channels, the voltage-gated Na^+ channel and the voltage-gated K^+ channel. The voltage-gated Na^+ channel actually has two gates. One is the **activation gate**, which opens when the membrane potential crosses -55 mV. The other gate is the **inactivation gate**, which closes after a specific period of time—on the order of a fraction of a millisecond of the activation opening. When a cell is at rest, the activation gate is closed and the inactivation gate is open. This is called the "closed but capable of opening configuration." When the threshold is reached, the activation gate opens, allowing Na^+ to rush into the cell. The voltage-gated Na+ channels are now in the “open” configuration as both activation and inactivation gates are open. Remember, the opening of the activation gate will trigger the slow closing of the inactivation gates. When the inactivation gate is finally closed and the activation gate is still open, the voltage-gated Na+ channel is now in the “closed and incapable of opening” configuration.

The voltage-gated K^+ channel has only one gate, which is sensitive to a membrane voltage of -50 mV. However, it does not open as quickly as the voltage-gated Na^+ channel does. It might take a fraction of a millisecond for the channel to open once that voltage has been reached. The timing of this coincides with the Na^+ channels are being inactivated. As the membrane potential repolarizes and the voltage passes -70 mV again, the channel closes—again, with a little delay. Potassium continues to leave the cell for a short while and the membrane potential becomes more negative. The voltage-gated K+ channels begin to close, but since they are slow this results in the hyperpolarizing. Once the channels are closed, the membrane can return to the resting membrane potential because of the ongoing activity of the leak channels and the Na^+/K^+ pump.

All of this takes place within approximately 2 milliseconds ([link]). While an action potential is in progress, the neuron is in the **refractory period**. There are two phases of the refractory period: the **absolute refractory** period and the **relative refractory period**. During the absolute phase, another action potential will not start. This is because of the inactivation gate of the voltage-gated Na^+ channel. Once that channel is back to its resting conformation (less than -55 mV), a new action potential could be started, but only by a stronger stimulus than the one that initiated the current action potential. This is because of the flow of K^+ out of the cell. Because that ion is rushing out, any Na+ that tries to enter will not depolarize the cell, but will only keep the cell from hyperpolarizing.

Stages of an Action Potential



Plotting voltage measured across the cell membrane against time, the events of the action potential can be related to specific changes in the membrane voltage. (1) At rest, the membrane voltage is -70 mV. (2) The membrane begins to depolarize when an external stimulus is applied. (3) The membrane voltage begins a rapid rise toward +30 mV. (4) The membrane voltage starts to return to a negative value. (5) Repolarization continues past the resting membrane voltage, resulting in hyperpolarization. (6) The membrane

voltage returns to the resting value shortly after hyperpolarization.

Propagation of the Action Potential

The action potential is initiated at the beginning of the axon, at what is called the axon hillock. There is a high density of voltage-gated Na^+ channels so that rapid depolarization can take place here. The action potential then travels along the axon through propagation. In unmyelinated axons, the action potential is propagated because more voltage-gated Na+ channels are opened as the depolarization spreads. This spreading occurs because Na^+ enters through the channel and moves along the inside of the cell membrane. As the Na^+ moves, or flows, a short distance along the cell membrane, its positive charge depolarizes a little more of the cell membrane. As that depolarization spreads, new voltage-gated Na^+ channels open and more ions rush into the cell, spreading the depolarization a little farther.

When myelination is present, the action potential propagates differently. Sodium ions that enter the cell at the initial segment start to spread along the length of the axon segment, but there are no voltage-gated Na^+ channels until the first node of Ranvier. Because there is not constant opening of these channels along the axon segment, the depolarization spreads at an optimal speed. The distance between nodes is the optimal distance to keep the membrane still depolarized above threshold at the next node. As Na^+ spreads along the inside of the membrane of the axon segment, the charge starts to dissipate. If the node were any farther down the axon, that depolarization would have fallen off too much for voltage-gated Na^+ channels to be activated at the next node of Ranvier. If the nodes were any closer together, the speed of propagation would be slower.

Propagation along an unmyelinated axon is referred to as **continuous conduction**; along the length of a myelinated axon, it is **saltatory conduction**. Continuous conduction is slow because there are always voltage-gated Na^+ channels opening, and more and more Na^+ is rushing into the cell. Saltatory conduction is faster because the action potential

basically jumps from one node to the next (saltare = “to leap”), and the new influx of Na^+ renews the depolarized membrane. Along with the myelination of the axon, the diameter of the axon can influence the speed of conduction. Much as water runs faster in a wide river than in a narrow creek, Na^+-based depolarization spreads faster down a wide axon than down a narrow one. This concept is known as **resistance** and is generally true for electrical wires or plumbing, just as it is true for axons, although the specific conditions are different at the scales of electrons or ions versus water in a river.

Note:
Homeostatic Imbalances
Potassium Concentration
Glial cells, especially astrocytes, are responsible for maintaining the chemical environment of the CNS tissue. The concentrations of ions in the extracellular fluid are the basis for how the membrane potential is established and changes in electrochemical signaling. If the balance of ions is upset, drastic outcomes are possible.
Normally the concentration of K^+ is higher inside the neuron than outside. After the repolarizing phase of the action potential, K^+ leakage channels and the Na^+/K^+ pump ensure that the ions return to their original locations. Following a stroke or other ischemic event, extracellular K^+ levels are elevated. The astrocytes in the area are equipped to clear excess K^+ to aid the pump. But when the level is far out of balance, the effects can be irreversible.
Astrocytes can become reactive in cases such as these, which impairs their ability to maintain the local chemical environment. The glial cells enlarge and their processes swell. They lose their K^+ buffering ability and the function of the pump is affected, or even reversed. If a Na^+ gradient breaks down, this has a more important effect than interrupting the action potential. Glucose transport into cells is coupled with Na^+ co-transport. When that is lost, the cell cannot get the energy it needs. In the central nervous system, carbohydrate metabolism is the only means of producing ATP. Elsewhere in the body, cells rely on carbohydrates, lipids, or amino acids to power mitochondrial ATP production. But the CNS does not store

lipids in adipocytes (fat cells) as an energy reserve. The lipids in the CNS are in the cell membranes of neurons and glial cells, notably as an integral component of myelin. Proteins in the CNS are crucial to neuronal function, in roles such as channels for electrical signaling or as part of the cytoskeleton. Those macromolecules are not used to power mitochondrial ATP production in neurons.

Note:



Visit this site to see a virtual neurophysiology lab, and to observe electrophysiological processes in the nervous system, where scientists directly measure the electrical signals produced by neurons. Often, the action potentials occur so rapidly that watching a screen to see them occur is not helpful. A speaker is powered by the signals recorded from a neuron and it “pops” each time the neuron fires an action potential. These action potentials are firing so fast that it sounds like static on the radio. Electrophysiologists can recognize the patterns within that static to understand what is happening. Why is the leech model used for measuring the electrical activity of neurons instead of using humans?

Chapter Review

The nervous system is characterized by electrical signals that are sent from one area to another. Whether those areas are close or very far apart, the signal must travel along an axon. The basis of the electrical signal is the controlled distribution of ions across the membrane. Transmembrane ion channels regulate when ions can move in or out of the cell, so that a precise

signal is generated. This signal is the action potential which has a very characteristic shape based on voltage changes across the membrane in a given time period.

The membrane is normally at rest with established Na^+ and K^+ concentrations on either side. A stimulus will start the depolarization of the membrane, and voltage-gated channels will result in further depolarization followed by repolarization of the membrane. A slight overshoot of hyperpolarization marks the end of the action potential. While an action potential is in progress, another cannot be generated under the same conditions. While the voltage-gated Na^+ channel is inactivated, absolutely no action potentials can be generated. Once that channel has returned to its resting state, a new action potential is possible, but it must be started by a relatively stronger stimulus to overcome the K^+ leaving the cell.

The action potential travels down the axon as voltage-gated ion channels are opened by the spreading depolarization. In unmyelinated axons, this happens in a continuous fashion because there are voltage-gated channels throughout the membrane. In myelinated axons, propagation is described as saltatory because voltage-gated channels are only found at the nodes of Ranvier and the electrical events seem to “jump” from one node to the next. Saltatory conduction is faster than continuous conduction, meaning that myelinated axons propagate their signals faster. The diameter of the axon also makes a difference as ions diffusing within the cell have less resistance in a wider space.

Glossary

absolute refractory period
: time during an action period when another action potential cannot be generated because the voltage-gated Na^+ channel is inactivated

activation gate
: part of the voltage-gated Na^+ channel that opens when the membrane voltage reaches threshold

continuous conduction

slow propagation of an action potential along an unmyelinated axon owing to voltage-gated Na^+ channels located along the entire length of the cell membrane

depolarization
: change in a cell membrane potential from rest toward zero

excitable membrane
: cell membrane that regulates the movement of ions so that an electrical signal can be generated

gated
: property of a channel that determines how it opens under specific conditions, such as voltage change or physical deformation

inactivation gate
: part of a voltage-gated Na^+ channel that closes when the membrane potential reaches +30 mV

leak channel
: ion channel that opens randomly and is not gated to a specific event, also known as a non-gated channel

ligand-gated channels
: another name for an ionotropic receptor for which a neurotransmitter is the ligand

mechanically gated channel
: ion channel that opens when a physical event directly affects the structure of the protein

membrane potential
: distribution of charge across the cell membrane, based on the charges of ions

refractory period
: time after the initiation of an action potential when another action potential cannot be generated

relative refractory period
: time during the refractory period when a new action potential can only be initiated by a stronger stimulus than the current action potential because voltage-gated K^+ channels are not closed

repolarization
: return of the membrane potential to its normally negative voltage at the end of the action potential

resistance
: property of an axon that relates to the ability of particles to diffuse through the cytoplasm; this is inversely proportional to the fiber diameter

resting membrane potential
: the difference in voltage measured across a cell membrane under steady-state conditions, typically -70 mV

saltatory conduction
: quick propagation of the action potential along a myelinated axon owing to voltage-gated Na^+ channels being present only at the nodes of Ranvier

voltage-gated channel
: ion channel that opens because of a change in the charge distributed across the membrane where it is located

Human Biology Chapter 8.3: Cardiac Muscle and Electrical Activity
By the end of this section, you will be able to:

- Describe the structure of cardiac muscle
- Identify and describe the components of the conducting system that distributes electrical impulses through the heart
- Relate characteristics of an electrocardiogram to events in the cardiac cycle
- Identify blocks that can interrupt the cardiac cycle

Recall that cardiac muscle shares a few characteristics with both skeletal muscle and smooth muscle, but it has some unique properties of its own. Not the least of these exceptional properties is its ability to initiate an electrical potential at a fixed rate that spreads rapidly from cell to cell to trigger the contraction mechanism. This property is known as **autorhythmicity**. Neither smooth nor skeletal muscle can do this. Even though cardiac muscle has autorhythmicity, heart rate is modulated by the endocrine and nervous systems.

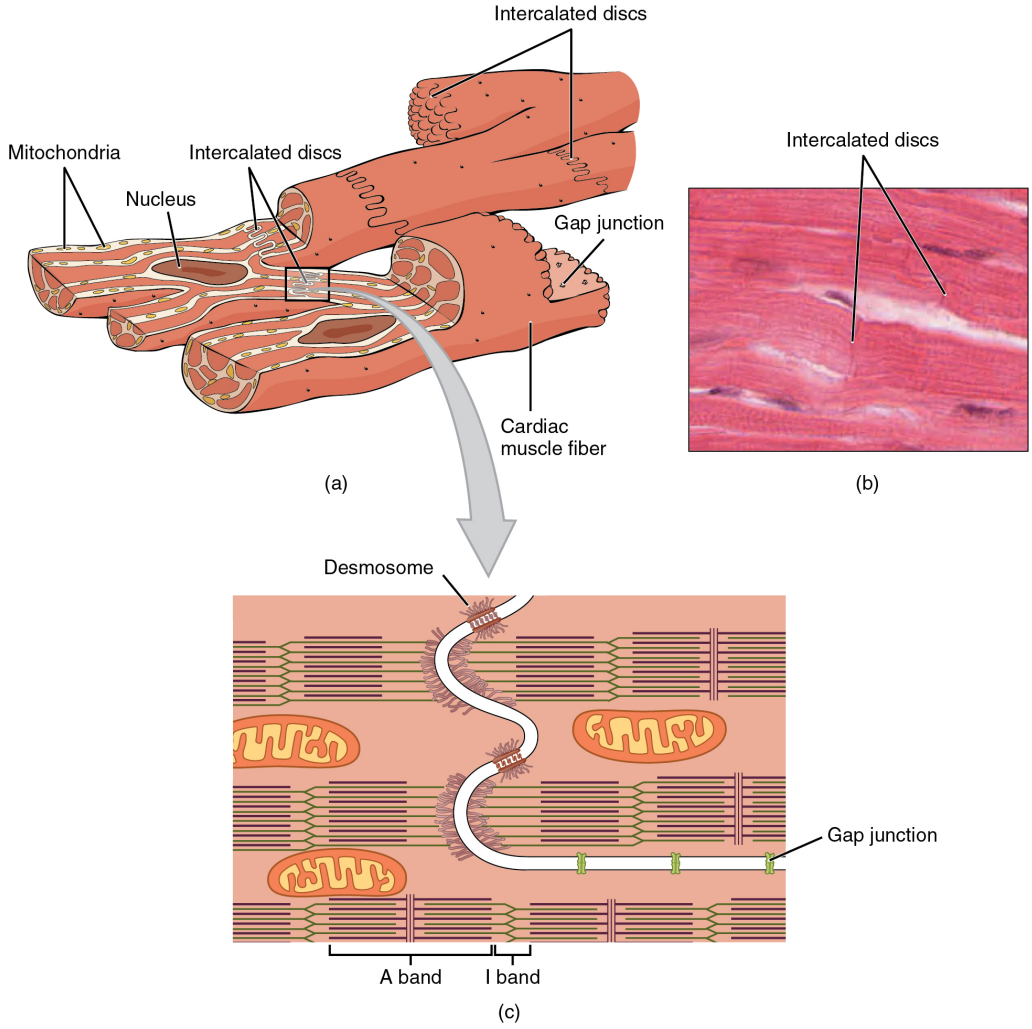
Structure of Cardiac Muscle

Compared to the giant cylinders of skeletal muscle, cardiac muscle cells, or cardiomyocytes, are considerably shorter with much smaller diameters. Cardiac muscle also demonstrates striations, the alternating pattern of dark A bands and light I bands attributed to the precise arrangement of the myofilaments and fibrils that are organized in sarcomeres along the length of the cell ([link]**a**). These contractile elements are virtually identical to skeletal muscle. T (transverse) tubules penetrate from the surface plasma membrane, the sarcolemma, to the interior of the cell, allowing the electrical impulse to reach the interior. The T tubules are only found at the Z discs, whereas in skeletal muscle, they are found at the junction of the A and I bands. Therefore, there are one-half as many T tubules in cardiac muscle as in skeletal muscle. In addition, the sarcoplasmic reticulum stores few calcium ions, so most of the calcium ions must come from outside the cells. The result is a slower onset of contraction. Mitochondria are plentiful, providing energy for the contractions of the heart. Typically,

cardiomyocytes have a single, central nucleus, but two or more nuclei may be found in some cells.

Cardiac muscle cells branch freely. A junction between two adjoining cells is marked by a critical structure called an **intercalated disc**, which helps support the synchronized contraction of the muscle ([link]**b**). The importance of strongly binding these cells together is necessitated by the forces exerted by contraction.

Cardiac Muscle



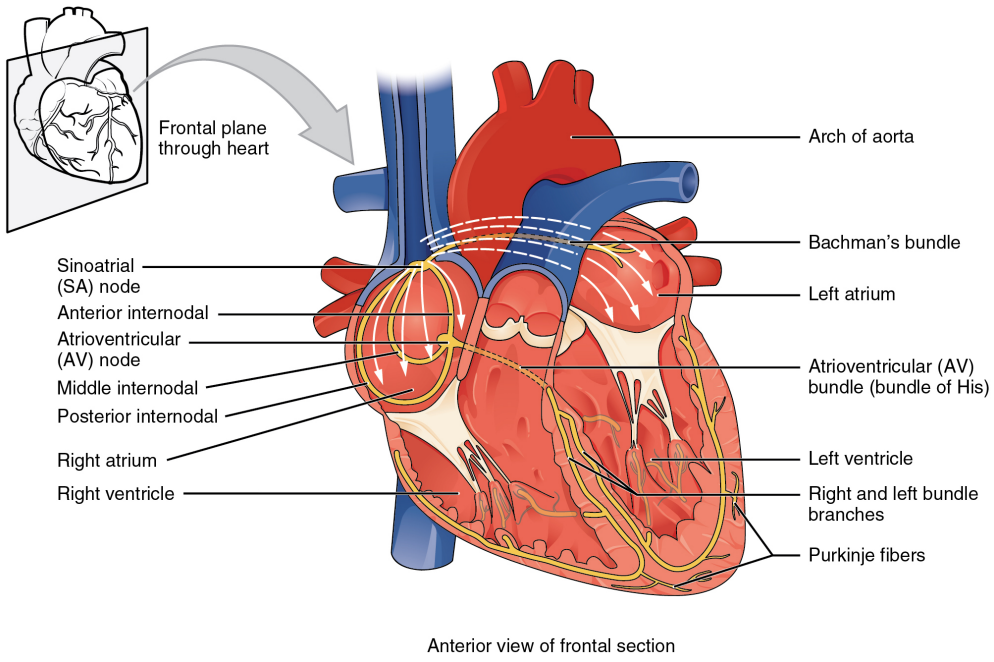
(a) Cardiac muscle cells have intercalated discs that are found at the junction of different cardiac muscle cells. (b) A photomicrograph of cardiac muscle cells shows the nuclei and intercalated discs. (Micrograph provided by the

Regents of the University of Michigan Medical School © 2012)

Note:
Everyday Connection
Repair and Replacement
Damaged cardiac muscle cells have extremely limited abilities to repair themselves or to replace dead cells via mitosis. Recent evidence indicates that at least some stem cells remain within the heart that continue to divide and at least potentially replace these dead cells. However, newly formed or repaired cells are rarely as functional as the original cells, and cardiac function is reduced. In the event of a heart attack or MI, dead cells are often replaced by patches of scar tissue. Autopsies performed on individuals who had successfully received heart transplants show some proliferation of original cells. If researchers can unlock the mechanism that generates new cells and restore full mitotic capabilities to heart muscle, the prognosis for heart attack survivors will be greatly enhanced.

Conduction System of the Heart

If embryonic heart cells are separated into a Petri dish and kept alive, each is capable of generating its own electrical impulse followed by contraction. When two independently beating embryonic cardiac muscle cells are placed together, the cell with the higher inherent rate sets the pace, and the impulse spreads from the faster to the slower cell to trigger a contraction. As more cells are joined together, the fastest cell continues to assume control of the rate. A fully developed adult heart maintains the capability of generating its own electrical impulse, triggered by the fastest cells, as part of the cardiac conduction system. The components of the cardiac conduction system include the sinoatrial (SA) node, the atrioventricular (AV) node, the atrioventricular bundle, the atrioventricular bundle branches, and the Purkinje fibers ([link]).
Conduction System of the Heart



Specialized conducting components of the heart include the sinoatrial (SA) node, the atrioventricular (AV) node, the atrioventricular bundle, the right and left bundle branches, and the Purkinje fibers.

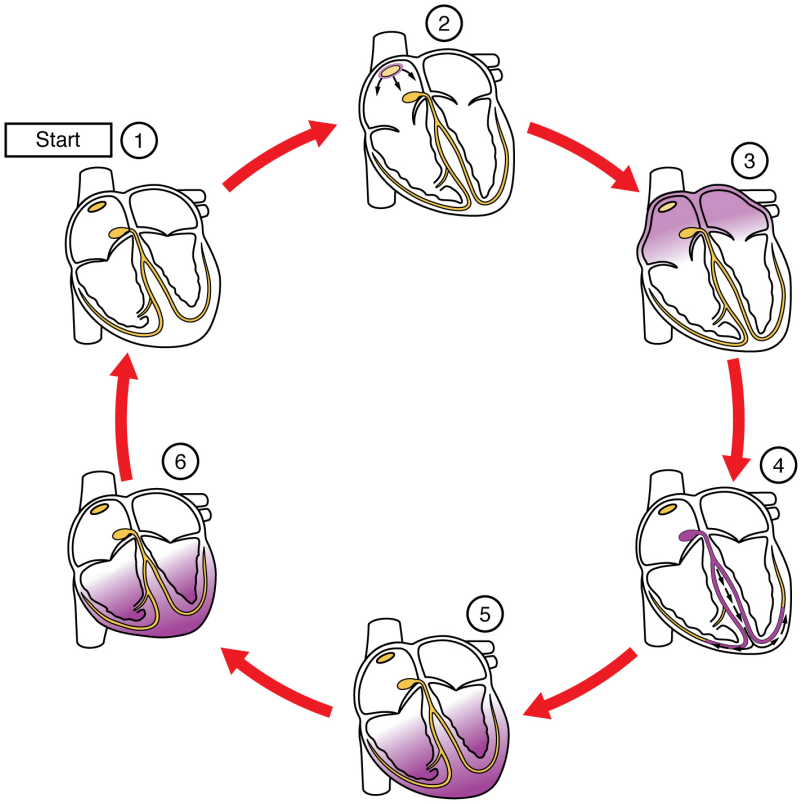
Sinoatrial (SA) Node

Normal cardiac rhythm is established by the **sinoatrial (SA) node**, a specialized clump of myocardial conducting cells located in the superior and posterior walls of the right atrium in close proximity to the opening of the superior vena cava. The SA node is known as the **pacemaker** of the heart. It initiates the **sinus rhythm**, or normal electrical pattern followed by contraction of the heart.

This impulse spreads from its initiation in the SA node throughout the atria to the atrioventricular (AV) node. The impulse takes approximately 50 ms (milliseconds) to travel between these two nodes. The connective tissue of the cardiac skeleton prevents the impulse from spreading into the

myocardial cells in the ventricles except at the atrioventricular node. [link] illustrates the initiation of the impulse in the SA node that then spreads the impulse throughout the atria to the atrioventricular node.

Cardiac Conduction



(1) The sinoatrial (SA) node and the remainder of the conduction system are at rest. (2) The SA node initiates the action potential, which sweeps across the atria. (3) After reaching the atrioventricular node, there is a delay of approximately 100 ms that allows the atria to complete pumping blood before the impulse is transmitted to the atrioventricular bundle. (4) Following the delay, the impulse travels through the atrioventricular bundle and bundle branches to the Purkinje fibers. (5) The impulse spreads to the contractile fibers of the ventricle. (6) Ventricular contraction begins.

Atrioventricular (AV) Node

The **atrioventricular (AV) node** is a second clump of specialized conductive cells, located in the lower portion of the right atrium within the atrioventricular wall. There is a critical pause before the AV node initiates an impulse and transmits it to the atrioventricular bundle (see [link], step 3).

Atrioventricular Bundle (Bundle of His), Bundle Branches, and Purkinje Fibers

Arising from the AV node, the **atrioventricular bundle**, proceeds through the septum before dividing into two **atrioventricular bundle branches**, commonly called the left and right bundle branches. The left bundle branch supplies the left ventricle, and the right bundle branch the right ventricle. Both bundle branches descend and reach the apex of the heart where they connect with the Purkinje fibers (see [link], step 4).

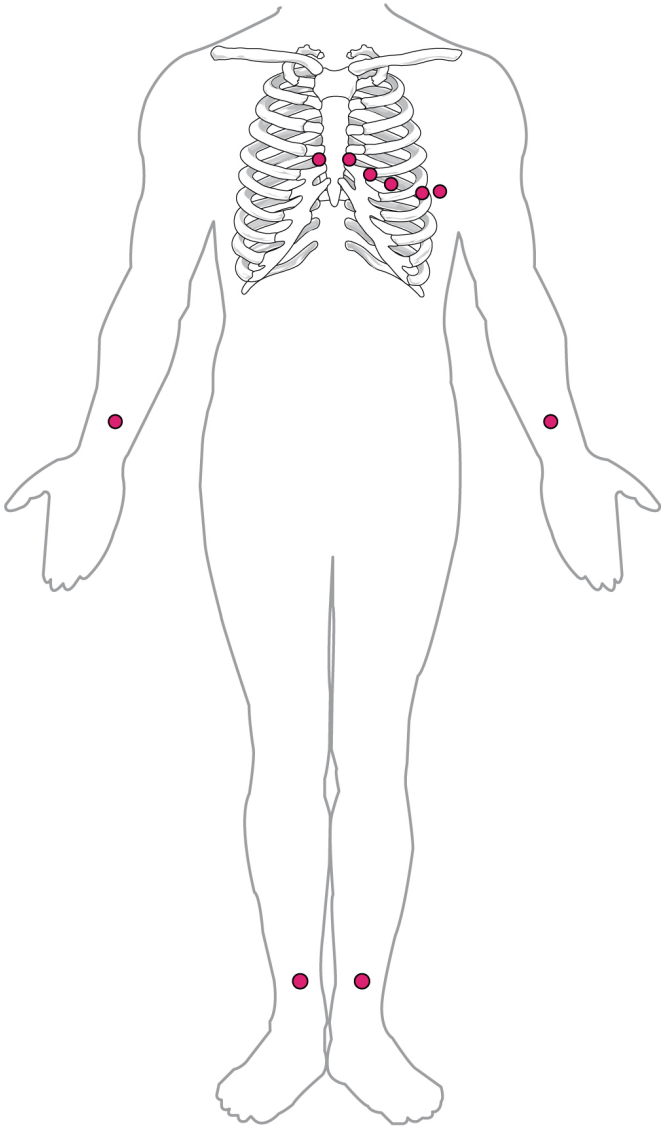
The **Purkinje fibers** are additional myocardial conductive fibers that spread the impulse to the myocardial contractile cells in the ventricles. Since the electrical stimulus begins at the apex, the contraction also begins at the apex and travels toward the base of the heart, similar to squeezing a tube of toothpaste from the bottom. This allows the blood to be pumped out of the ventricles and into the aorta and pulmonary trunk.

Electrocardiogram

By careful placement of surface electrodes on the body, it is possible to record the complex, compound electrical signal of the heart. This tracing of the electrical signal is the **electrocardiogram (ECG)**, also commonly abbreviated EKG (K coming kardiology, from the German term for cardiology). Careful analysis of the ECG reveals a detailed picture of both normal and abnormal heart function, and is an indispensable clinical

diagnostic tool. The standard electrocardiograph (the instrument that generates an ECG) uses 3, 5, or 12 leads. The greater the number of leads an electrocardiograph uses, the more information the ECG provides. The term “lead” may be used to refer to the cable from the electrode to the electrical recorder, but it typically describes the voltage difference between two of the electrodes. The 12-lead electrocardiograph uses 10 electrodes placed in standard locations on the patient’s skin ([link]). In continuous ambulatory electrocardiographs, the patient wears a small, portable, battery-operated device known as a Holter monitor, that continuously monitors heart electrical activity, typically for a period of 24 hours during the patient’s normal routine.

Standard Placement of ECG Leads



In a 12-lead ECG, six electrodes are placed on the chest, and four electrodes are placed on the limbs.

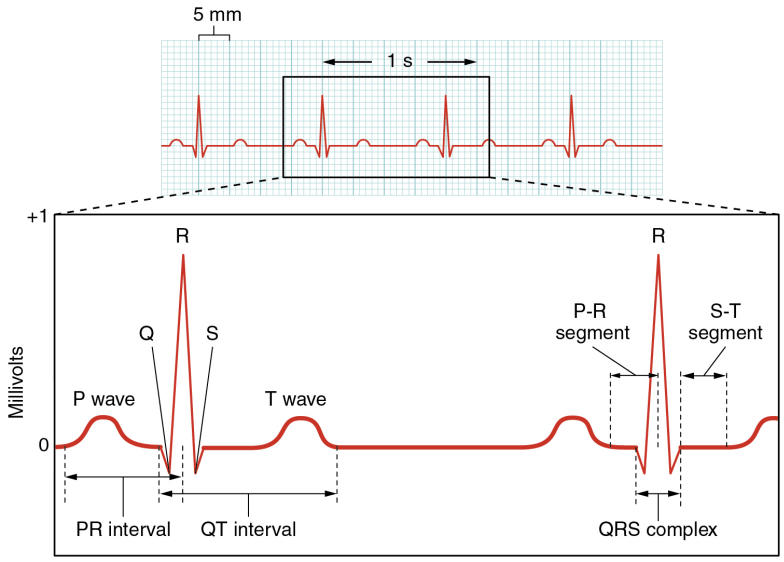
A normal ECG tracing is presented in [link]. Each component, segment, and interval is labeled and corresponds to important electrical events, demonstrating the relationship between these events and contraction in the heart.

There are five prominent points on the ECG: the P wave, the QRS complex, and the T wave. The small **P wave** represents the depolarization of the atria. The atria begin contracting approximately 25 ms after the start of the P wave. The large **QRS complex** represents the depolarization of the ventricles, which requires a much stronger electrical signal because of the larger size of the ventricular cardiac muscle. The ventricles begin to contract as the QRS reaches the peak of the R wave. Lastly, the **T wave** represents the repolarization of the ventricles. The repolarization of the atria occurs during the QRS complex, which masks it on an ECG.

Note:

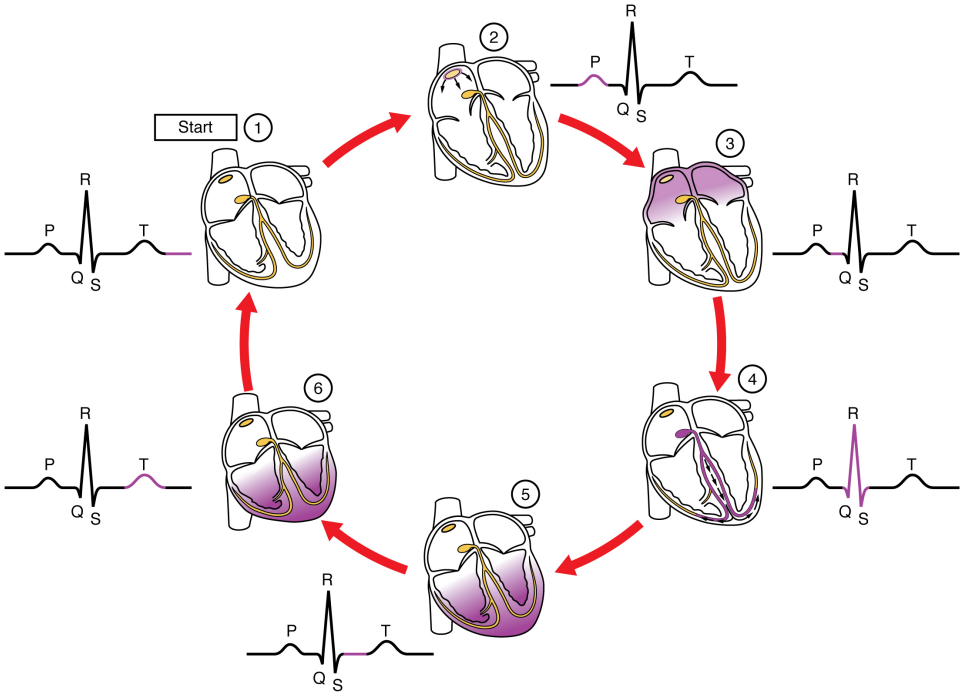
Visit this site for a more detailed analysis of ECGs.

Electrocardiogram



A normal tracing shows the P wave, QRS complex, and T wave. Additional segments and intervals are also shown.

ECG Tracing Correlated to the Cardiac Cycle



This diagram correlates an ECG tracing with the

electrical and mechanical events of a heart contraction. Each segment of an ECG tracing corresponds to one event in the cardiac cycle.

Note:

Everyday Connection

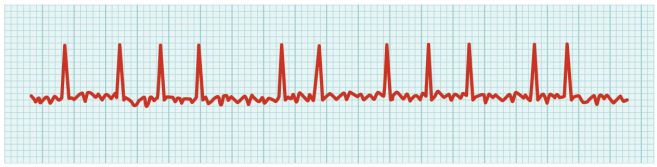
While interpretation of an ECG is possible and extremely valuable after some training, a full understanding of the complexities and intricacies generally requires several years of experience. In general, the size of the electrical variations, the duration of the events, and detailed analysis provide the most comprehensive picture of cardiac function. For example, an amplified P wave may indicate enlargement of the atria and an enlarged Q wave may indicate a MI (Myocardial Infarction). T waves often appear flatter when insufficient oxygen is being delivered to the myocardium. As useful as analyzing these electrical recordings may be, there are limitations. For example, not all areas suffering a MI may be obvious on the ECG. Additionally, it will not reveal the effectiveness of the pumping, which requires further testing, such as an ultrasound test called an echocardiogram or nuclear medicine imaging. Common abnormalities that may be detected by the ECGs are shown in [link].

Common ECG Abnormalities



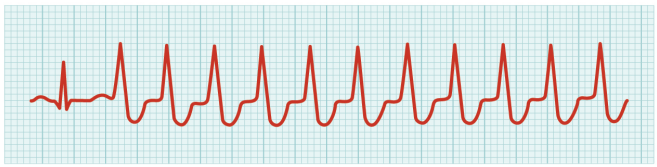
(a) Second-degree (partial) block

Note how half of the P waves are not followed by the QRS complex and T waves while the other half are.
Question: What would you expect to happen to heart rate (pulse)?



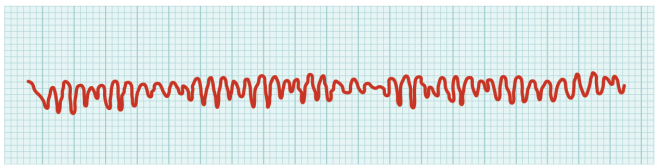
(b) Atrial fibrillation

Note the abnormal electrical pattern prior to the QRS complexes. Also note how the frequency between the QRS complexes has increased.
Question: What would you expect to happen to heart rate (pulse)?



(c) Ventricular tachycardia

Note the unusual shape of the QRS complex, focusing on the "S" component.
Question: What would you expect to happen to heart rate (pulse)?



(d) Ventricular fibrillation

Note the total lack of normal electrical activity.
Question: What would you expect to happen to heart rate (pulse)?



(e) Third-degree block

Note that in a third-degree block some of the impulses initiated by the SA node do not reach the AV node while others do. Also note that the P waves are not followed by the QRS complex.
Question: What would you expect to happen to heart rate (pulse)?

(a) In a second-degree or partial block, one-half of the P waves are not followed by the QRS complex and T waves while the other half are. (b) In atrial fibrillation, the electrical pattern is abnormal prior to the QRS complex, and the frequency between the QRS complexes has increased. (c) In ventricular tachycardia, the shape of the QRS complex is abnormal. (d) In ventricular fibrillation, there is no normal electrical activity. (e) In a third-degree block, there is no correlation between atrial activity (the P wave) and ventricular activity (the QRS complex).

Note:

Everyday Connection

External Automated Defibrillators

In the event that the electrical activity of the heart is severely disrupted, cessation of electrical activity or fibrillation may occur. In fibrillation, the heart beats in a wild, uncontrolled manner, which prevents it from being able to pump effectively. Atrial fibrillation (see [link]**b**) is a serious condition, but as long as the ventricles continue to pump blood, the patient's life may not be in immediate danger. Ventricular fibrillation (see [link]**d**) is a medical emergency that requires life support, because the ventricles are not effectively pumping blood. In a hospital setting, it is often described as "code blue." If untreated for as little as a few minutes, ventricular fibrillation may lead to brain death. The most common treatment is defibrillation, which uses special paddles to apply a charge to the heart from an external electrical source in an attempt to establish a normal sinus rhythm ([link]). A defibrillator effectively stops the heart so that the SA node can trigger a normal conduction cycle. Because of their effectiveness in reestablishing a normal sinus rhythm, external automated defibrillators (EADs) are being placed in areas frequented by large numbers of people, such as schools, restaurants, and airports. These devices contain simple and direct verbal instructions that can be followed by nonmedical personnel in an attempt to save a life.

Defibrillators



(a) An external automatic defibrillator can be used by nonmedical personnel to reestablish a normal sinus rhythm in a person with fibrillation. (b) Defibrillator paddles are

more commonly used in hospital settings. (credit b: “widerider107”/flickr.com)

When arrhythmias become a chronic problem, the heart maintains a junctional rhythm, which originates in the AV node. In order to speed up the heart rate and restore full sinus rhythm, a cardiologist can implant an **artificial pacemaker**, which delivers electrical impulses to the heart muscle to ensure that the heart continues to contract and pump blood effectively. These artificial pacemakers are programmable by the cardiologists and can either provide stimulation temporarily upon demand or on a continuous basis. Some devices also contain built-in defibrillators.

Chapter Review

The heart is regulated by both neural and endocrine (i.e. hormonal) control, yet it is capable of initiating its own action potential followed by muscular contraction. The conductive cells within the heart establish the heart rate and transmit it through the myocardium. The contractile cells contract and propel the blood. The normal path of transmission for the conductive cells is the sinoatrial (SA) node, atrioventricular (AV) node, atrioventricular (AV) bundle, bundle branches, and Purkinje fibers. Recognizable points on the ECG include the P wave that corresponds to atrial depolarization (i.e. contraction), the QRS complex that corresponds to ventricular depolarization, and the T wave that corresponds to ventricular repolarization (relaxation).

Review Questions

Exercise:

Problem: Which of the following is unique to cardiac muscle cells?

a. Only cardiac muscle contains a sarcoplasmic reticulum.
b. Only cardiac muscle has gap junctions.
c. Only cardiac muscle is capable of autorhythmicity
d. Only cardiac muscle has a high concentration of mitochondria.

Solution:

C

Exercise:

Problem:

Which portion of the ECG corresponds to repolarization (i.e. relaxation) of the atria?

a. P wave
b. QRS complex
c. T wave
d. none of the above: atrial repolarization is masked by ventricular depolarization

---

Solution:

D

Critical Thinking Questions

Exercise:

Problem:

How does the delay of the impulse at the atrioventricular node contribute to cardiac function?

Solution:

It ensures sufficient time for the atrial muscle to contract and pump blood into the ventricles prior to the impulse being conducted into the lower chambers.

Glossary

artificial pacemaker
: medical device that transmits electrical signals to the heart to ensure that it contracts and pumps blood to the body

atrioventricular bundle
: (also, bundle of His) group of specialized myocardial conductile cells that transmit the impulse from the AV node through the interventricular septum; form the left and right atrioventricular bundle branches

atrioventricular bundle branches
: (also, left or right bundle branches) specialized myocardial conductile cells that arise from the bifurcation of the atrioventricular bundle and pass through the interventricular septum; lead to the Purkinje fibers and also to the right papillary muscle via the moderator band

atrioventricular (AV) node
: clump of myocardial cells located in the inferior portion of the right atrium within the atrioventricular septum; receives the impulse from the SA node, pauses, and then transmits it into specialized conducting cells within the interventricular septum

autorhythmicity
: ability of cardiac muscle to initiate its own electrical impulse that triggers the mechanical contraction that pumps blood at a fixed pace without nervous or endocrine control

Bachmann’s bundle
: (also, interatrial band) group of specialized conducting cells that transmit the impulse directly from the SA node in the right atrium to the left atrium

bundle of His
: (also, atrioventricular bundle) group of specialized myocardial conductile cells that transmit the impulse from the AV node through the interventricular septum; form the left and right atrioventricular bundle branches

electrocardiogram (ECG)
: surface recording of the electrical activity of the heart that can be used for diagnosis of irregular heart function; also abbreviated as EKG

heart block
: interruption in the normal conduction pathway

interatrial band
: (also, Bachmann's bundle) group of specialized conducting cells that transmit the impulse directly from the SA node in the right atrium to the left atrium

intercalated disc
: physical junction between adjacent cardiac muscle cells; consisting of desmosomes, specialized linking proteoglycans, and gap junctions that allow passage of ions between the two cells

internodal pathways
: specialized conductile cells within the atria that transmit the impulse from the SA node throughout the myocardial cells of the atrium and to the AV node

myocardial conducting cells
: specialized cells that transmit electrical impulses throughout the heart and trigger contraction by the myocardial contractile cells

myocardial contractile cells
: bulk of the cardiac muscle cells in the atria and ventricles that conduct impulses and contract to propel blood

P wave

component of the electrocardiogram that represents the depolarization of the atria

pacemaker
: cluster of specialized myocardial cells known as the SA node that initiates the sinus rhythm

prepotential depolarization
: (also, spontaneous depolarization) mechanism that accounts for the autorhythmic property of cardiac muscle; the membrane potential increases as sodium ions diffuse through the always-open sodium ion channels and causes the electrical potential to rise

Purkinje fibers
: specialized myocardial conduction fibers that arise from the bundle branches and spread the impulse to the myocardial contraction fibers of the ventricles

QRS complex
: component of the electrocardiogram that represents the depolarization of the ventricles and includes, as a component, the repolarization of the atria

sinoatrial (SA) node
: known as the pacemaker, a specialized clump of myocardial conducting cells located in the superior portion of the right atrium that has the highest inherent rate of depolarization that then spreads throughout the heart

sinus rhythm
: normal contractile pattern of the heart

spontaneous depolarization
: (also, prepotential depolarization) the mechanism that accounts for the autorhythmic property of cardiac muscle; the membrane potential increases as sodium ions diffuse through the always-open sodium ion channels and causes the electrical potential to rise

T wave
: component of the electrocardiogram that represents the repolarization of the ventricles

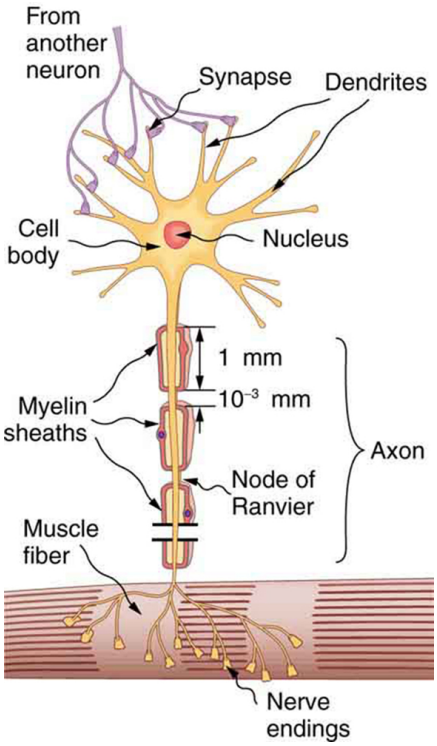
Nerve Conduction–Electrocardiograms

- Explain the process by which electric signals are transmitted along a neuron.
- Explain the effects myelin sheaths have on signal propagation.
- Explain what the features of an ECG signal indicate.

Nerve Conduction

Electric currents in the vastly complex system of billions of nerves in our body allow us to sense the world, control parts of our body, and think. These are representative of the three major functions of nerves. First, nerves carry messages from our sensory organs and others to the central nervous system, consisting of the brain and spinal cord. Second, nerves carry messages from the central nervous system to muscles and other organs. Third, nerves transmit and process signals within the central nervous system. The sheer number of nerve cells and the incredibly greater number of connections between them makes this system the subtle wonder that it is. **Nerve conduction** is a general term for electrical signals carried by nerve cells. It is one aspect of **bioelectricity**, or electrical effects in and created by biological systems.

Nerve cells, properly called *neurons*, look different from other cells—they have tendrils, some of them many centimeters long, connecting them with other cells. (See [link].) Signals arrive at the cell body across *synapses* or through *dendrites*, stimulating the neuron to generate its own signal, sent along its long *axon* to other nerve or muscle cells. Signals may arrive from many other locations and be transmitted to yet others, conditioning the synapses by use, giving the system its complexity and its ability to learn.

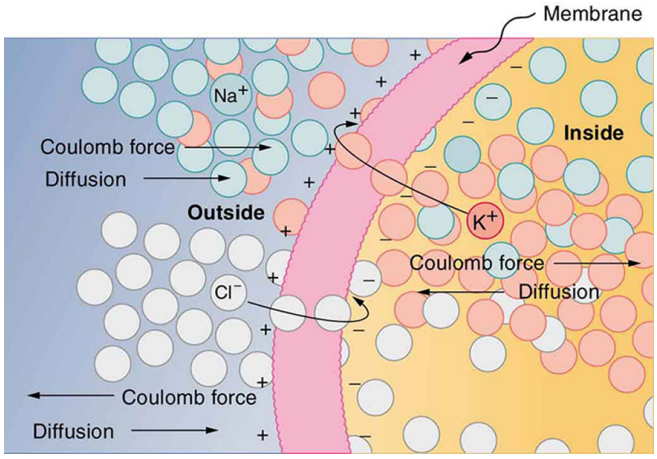


A neuron with its dendrites and long axon. Signals in the form of electric currents reach the cell body through dendrites and across synapses, stimulating the neuron to generate its own signal sent down the axon. The number of interconnections can be far greater than shown here.

The method by which these electric currents are generated and transmitted is more complex than the simple movement of free charges in a conductor,

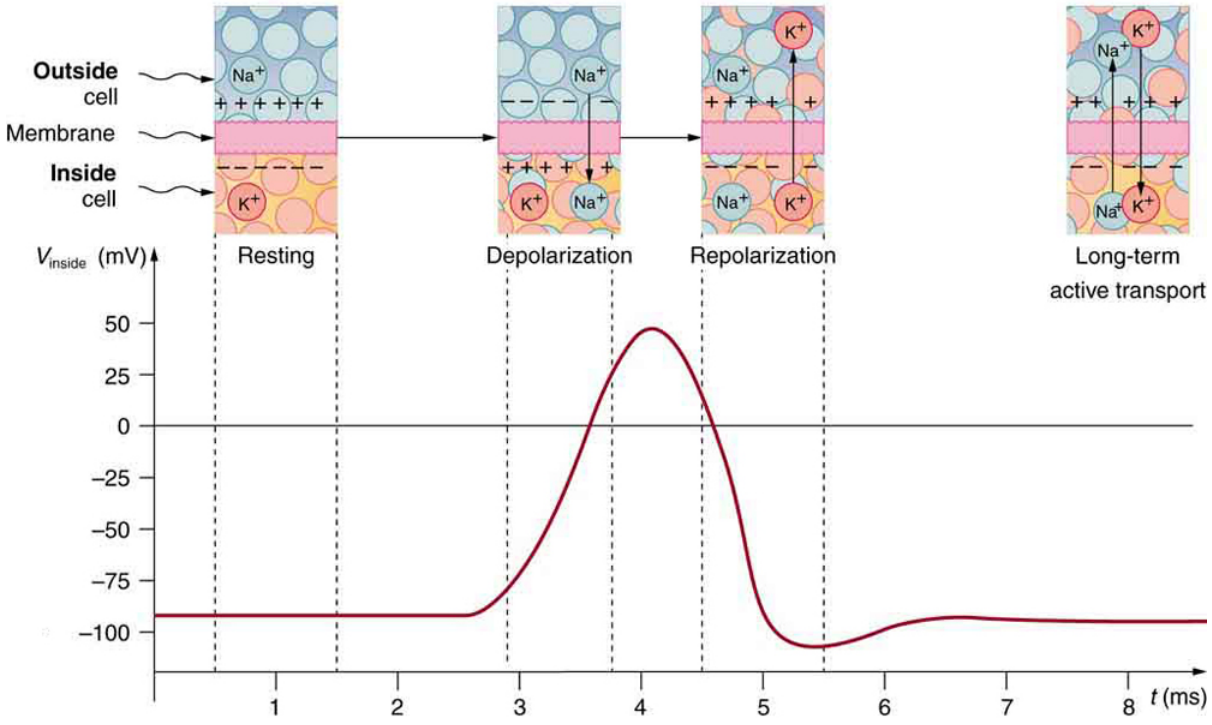
but it can be understood with principles already discussed in this text. The most important of these are the Coulomb force and diffusion.

[link] illustrates how a voltage (potential difference) is created across the cell membrane of a neuron in its resting state. This thin membrane separates electrically neutral fluids having differing concentrations of ions, the most important varieties being Na^+, K^+, and Cl^- (these are sodium, potassium, and chlorine ions with single plus or minus charges as indicated). As discussed in Molecular Transport Phenomena: Diffusion, Osmosis, and Related Processes, free ions will diffuse from a region of high concentration to one of low concentration. But the cell membrane is **semipermeable**, meaning that some ions may cross it while others cannot. In its resting state, the cell membrane is permeable to K^+ and Cl^-, and impermeable to Na^+. Diffusion of K^+ and Cl^- thus creates the layers of positive and negative charge on the outside and inside of the membrane. The Coulomb force prevents the ions from diffusing across in their entirety. Once the charge layer has built up, the repulsion of like charges prevents more from moving across, and the attraction of unlike charges prevents more from leaving either side. The result is two layers of charge right on the membrane, with diffusion being balanced by the Coulomb force. A tiny fraction of the charges move across and the fluids remain neutral (other ions are present), while a separation of charge and a voltage have been created across the membrane.



The semipermeable membrane of a

cell has different concentrations of ions inside and out. Diffusion moves the K^+ and Cl^- ions in the direction shown, until the Coulomb force halts further transfer. This results in a layer of positive charge on the outside, a layer of negative charge on the inside, and thus a voltage across the cell membrane. The membrane is normally impermeable to Na^+.



An action potential is the pulse of voltage inside a nerve cell graphed here. It is caused by movements of ions across the cell membrane as shown. Depolarization occurs when a stimulus makes the membrane permeable to Na^+ ions. Repolarization follows as the membrane

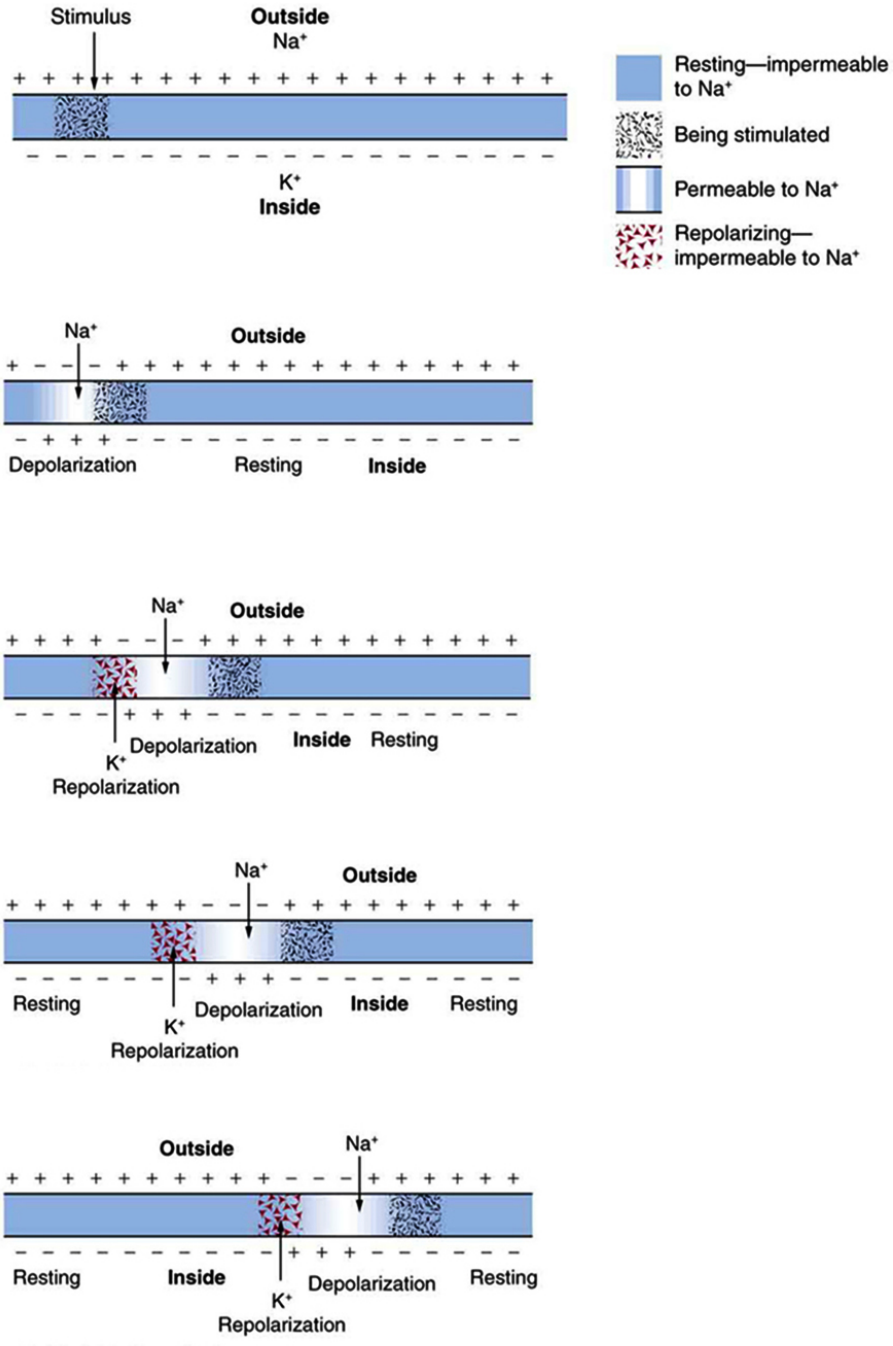
again becomes impermeable to Na^+, and K^+ moves from high to low concentration. In the long term, active transport slowly maintains the concentration differences, but the cell may fire hundreds of times in rapid succession without seriously depleting them.

The separation of charge creates a potential difference of 70 to 90 mV across the cell membrane. While this is a small voltage, the resulting electric field ($E = V/d$) across the only 8-nm-thick membrane is immense (on the order of 11 MV/m!) and has fundamental effects on its structure and permeability. Now, if the exterior of a neuron is taken to be at 0 V, then the interior has a *resting potential* of about –90 mV. Such voltages are created across the membranes of almost all types of animal cells but are largest in nerve and muscle cells. In fact, fully 25% of the energy used by cells goes toward creating and maintaining these potentials.

Electric currents along the cell membrane are created by any stimulus that changes the membrane's permeability. The membrane thus temporarily becomes permeable to Na^+, which then rushes in, driven both by diffusion and the Coulomb force. This inrush of Na^+ first neutralizes the inside membrane, or *depolarizes* it, and then makes it slightly positive. The depolarization causes the membrane to again become impermeable to Na^+, and the movement of K^+ quickly returns the cell to its resting potential, or *repolarizes* it. This sequence of events results in a voltage pulse, called the *action potential*. (See [link].) Only small fractions of the ions move, so that the cell can fire many hundreds of times without depleting the excess concentrations of Na^+ and K^+. Eventually, the cell must replenish these ions to maintain the concentration differences that create bioelectricity. This sodium-potassium pump is an example of *active transport*, wherein cell energy is used to move ions across membranes against diffusion gradients and the Coulomb force.

The action potential is a voltage pulse at one location on a cell membrane. How does it get transmitted along the cell membrane, and in particular down an axon, as a nerve impulse? The answer is that the changing voltage and electric fields affect the permeability of the adjacent cell membrane, so

that the same process takes place there. The adjacent membrane depolarizes, affecting the membrane further down, and so on, as illustrated in [link]. Thus the action potential stimulated at one location triggers a *nerve impulse* that moves slowly (about 1 m/s) along the cell membrane.



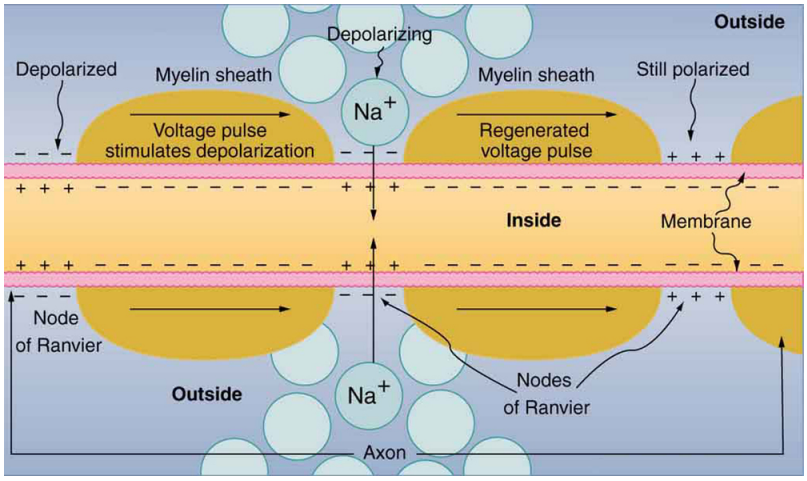
A nerve impulse is the propagation of an action potential along a cell membrane. A stimulus causes an action potential at one location, which changes the permeability of the adjacent membrane, causing an action potential there. This in turn affects the membrane further down, so that the action potential moves slowly (in electrical terms) along the cell membrane. Although the impulse is due to Na^+ and K^+ going across the membrane, it is equivalent to a wave of charge moving along the outside and inside of the membrane.

Some axons, like that in [link], are sheathed with *myelin*, consisting of fat-containing cells. [link] shows an enlarged view of an axon having myelin sheaths characteristically separated by unmyelinated gaps (called nodes of Ranvier). This arrangement gives the axon a number of interesting properties. Since myelin is an insulator, it prevents signals from jumping between adjacent nerves (cross talk). Additionally, the myelinated regions transmit electrical signals at a very high speed, as an ordinary conductor or resistor would. There is no action potential in the myelinated regions, so that no cell energy is used in them. There is an IR signal loss in the myelin, but the signal is regenerated in the gaps, where the voltage pulse triggers the action potential at full voltage. So a myelinated axon transmits a nerve impulse faster, with less energy consumption, and is better protected from cross talk than an unmyelinated one. Not all axons are myelinated, so that cross talk and slow signal transmission are a characteristic of the normal operation of these axons, another variable in the nervous system.

The degeneration or destruction of the myelin sheaths that surround the nerve fibers impairs signal transmission and can lead to numerous neurological effects. One of the most prominent of these diseases comes from the body’s own immune system attacking the myelin in the central nervous system—multiple sclerosis. MS symptoms include fatigue, vision problems, weakness of arms and legs, loss of balance, and tingling or

numbness in one’s extremities (neuropathy). It is more apt to strike younger adults, especially females. Causes might come from infection, environmental or geographic affects, or genetics. At the moment there is no known cure for MS.

Most animal cells can fire or create their own action potential. Muscle cells contract when they fire and are often induced to do so by a nerve impulse. In fact, nerve and muscle cells are physiologically similar, and there are even hybrid cells, such as in the heart, that have characteristics of both nerves and muscles. Some animals, like the infamous electric eel (see [link]), use muscles ganged so that their voltages add in order to create a shock great enough to stun prey.



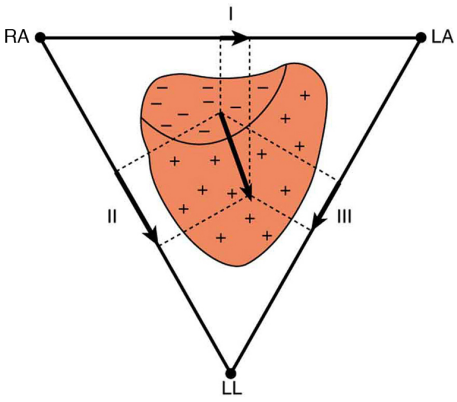
Propagation of a nerve impulse down a myelinated axon, from left to right. The signal travels very fast and without energy input in the myelinated regions, but it loses voltage. It is regenerated in the gaps. The signal moves faster than in unmyelinated axons and is insulated from signals in other nerves, limiting cross talk.



An electric eel flexes its muscles to create a voltage that stuns prey. (credit: chrisbb, Flickr)

Electrocardiograms

Just as nerve impulses are transmitted by depolarization and repolarization of adjacent membrane, the depolarization that causes muscle contraction can also stimulate adjacent muscle cells to depolarize (fire) and contract. Thus, a depolarization wave can be sent across the heart, coordinating its rhythmic contractions and enabling it to perform its vital function of propelling blood through the circulatory system. [link] is a simplified graphic of a depolarization wave spreading across the heart from the *sinoarterial (SA) node***,** the heart's natural pacemaker.

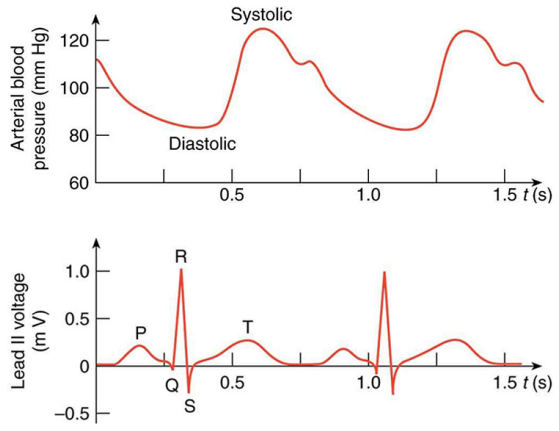


The outer surface of the heart changes from positive to negative during depolarization. This wave of depolarization is spreading from the top of the heart and is represented by a vector pointing in the direction of the wave. This vector is a voltage (potential difference) vector. Three electrodes, labeled RA, LA, and LL, are placed on the patient. Each pair (called leads I, II, and III) measures a component of the depolarization vector and is graphed in an ECG.

An **electrocardiogram (ECG)** is a record of the voltages created by the wave of depolarization and subsequent repolarization in the heart. Voltages between pairs of electrodes placed on the chest are vector components of the voltage wave on the heart. Standard ECGs have 12 or more electrodes, but only three are shown in [link] for clarity. Decades ago, three-electrode ECGs were performed by placing electrodes on the left and right arms and the left leg. The voltage between the right arm and the left leg is called the *lead II potential* and is the most often graphed. We shall examine the lead II potential as an indicator of heart-muscle function and see that it is coordinated with arterial blood pressure as well.

Heart function and its four-chamber action are explored in Viscosity and Laminar Flow; Poiseuille's Law. Basically, the right and left atria receive blood from the body and lungs, respectively, and pump the blood into the ventricles. The right and left ventricles, in turn, pump blood through the lungs and the rest of the body, respectively. Depolarization of the heart muscle causes it to contract. After contraction it is repolarized to ready it for the next beat. The ECG measures components of depolarization and repolarization of the heart muscle and can yield significant information on the functioning and malfunctioning of the heart.

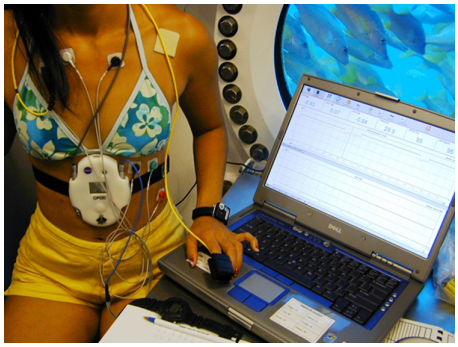
[link] shows an ECG of the lead II potential and a graph of the corresponding arterial blood pressure. The major features are labeled P, Q, R, S, and T. The *P wave* is generated by the depolarization and contraction of the atria as they pump blood into the ventricles. The *QRS complex* is created by the depolarization of the ventricles as they pump blood to the lungs and body. Since the shape of the heart and the path of the depolarization wave are not simple, the QRS complex has this typical shape and time span. The lead II QRS signal also masks the repolarization of the atria, which occur at the same time. Finally, the *T wave* is generated by the repolarization of the ventricles and is followed by the next P wave in the next heartbeat. Arterial blood pressure varies with each part of the heartbeat, with systolic (maximum) pressure occurring closely after the QRS complex, which signals contraction of the ventricles.



A lead II ECG with

corresponding arterial blood pressure. The QRS complex is created by the depolarization and contraction of the ventricles and is followed shortly by the maximum or systolic blood pressure. See text for further description.

Taken together, the 12 leads of a state-of-the-art ECG can yield a wealth of information about the heart. For example, regions of damaged heart tissue, called infarcts, reflect electrical waves and are apparent in one or more lead potentials. Subtle changes due to slight or gradual damage to the heart are most readily detected by comparing a recent ECG to an older one. This is particularly the case since individual heart shape, size, and orientation can cause variations in ECGs from one individual to another. ECG technology has advanced to the point where a portable ECG monitor with a liquid crystal instant display and a printer can be carried to patients' homes or used in emergency vehicles. See [link].



This NASA scientist and NEEMO 5 aquanaut’s heart rate and other vital signs

are being recorded by a portable device while living in an underwater habitat. (credit: NASA, Life Sciences Data Archive at Johnson Space Center, Houston, Texas)

Note:
PhET Explorations: Neuron
Stimulate a neuron and monitor what happens. Pause, rewind, and move forward in time in order to observe the ions as they move across the neuron membrane.
https://phet.colorado.edu/sims/html/neuron/latest/neuron_en.html

Section Summary

- Electric potentials in neurons and other cells are created by ionic concentration differences across semipermeable membranes.
- Stimuli change the permeability and create action potentials that propagate along neurons.
- Myelin sheaths speed this process and reduce the needed energy input.
- This process in the heart can be measured with an electrocardiogram (ECG).

Conceptual Questions

Exercise:

Problem:

Note that in [link], both the concentration gradient and the Coulomb force tend to move Na^+ ions into the cell. What prevents this?

Exercise:

Problem:

Define depolarization, repolarization, and the action potential.

Exercise:

Problem:

Explain the properties of myelinated nerves in terms of the insulating properties of myelin.

Problems & Exercises

Exercise:

Problem: **Integrated Concepts**

Use the ECG in [link] to determine the heart rate in beats per minute assuming a constant time between beats.

Solution:

80 beats/minute

Exercise:

Problem: **Integrated Concepts**

(a) Referring to [link], find the time systolic pressure lags behind the middle of the QRS complex. (b) Discuss the reasons for the time lag.

Glossary

nerve conduction
: the transport of electrical signals by nerve cells

bioelectricity
: electrical effects in and created by biological systems

semipermeable
: property of a membrane that allows only certain types of ions to cross it

electrocardiogram (ECG)
: usually abbreviated ECG, a record of voltages created by depolarization and repolarization, especially in the heart

How Neurons Communicate
By the end of this section, you will be able to:

- Describe the basis of the resting membrane potential
- Explain the stages of an action potential and how action potentials are propagated
- Explain the similarities and differences between chemical and electrical synapses
- Describe long-term potentiation and long-term depression

All functions performed by the nervous system—from a simple motor reflex to more advanced functions like making a memory or a decision—require neurons to communicate with one another. While humans use words and body language to communicate, neurons use electrical and chemical signals. Just like a person in a committee, one neuron usually receives and synthesizes messages from multiple other neurons before “making the decision” to send the message on to other neurons.

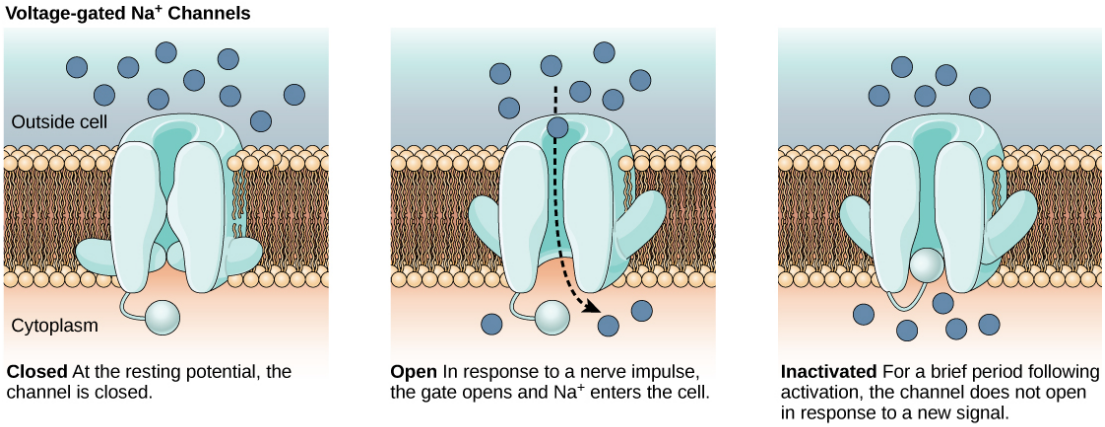
Nerve Impulse Transmission within a Neuron

For the nervous system to function, neurons must be able to send and receive signals. These signals are possible because each neuron has a charged cellular membrane (a voltage difference between the inside and the outside), and the charge of this membrane can change in response to neurotransmitter molecules released from other neurons and environmental stimuli. To understand how neurons communicate, one must first understand the basis of the baseline or ‘resting’ membrane charge.

Neuronal Charged Membranes

The lipid bilayer membrane that surrounds a neuron is impermeable to charged molecules or ions. To enter or exit the neuron, ions must pass through special proteins called ion channels that span the membrane. Ion channels have different configurations: open, closed, and inactive, as illustrated in [link]. Some ion channels need to be activated in order to open and allow ions to pass into or out of the cell. These ion channels are

sensitive to the environment and can change their shape accordingly. Ion channels that change their structure in response to voltage changes are called voltage-gated ion channels. Voltage-gated ion channels regulate the relative concentrations of different ions inside and outside the cell. The difference in total charge between the inside and outside of the cell is called the **membrane potential**.



Voltage-gated ion channels open in response to changes in membrane voltage. After activation, they become inactivated for a brief period and will no longer open in response to a signal.

Note:
Link to Learning

This video discusses the basis of the resting membrane potential.
https://www.openstaxcollege.org/l/resting_neuron

Resting Membrane Potential

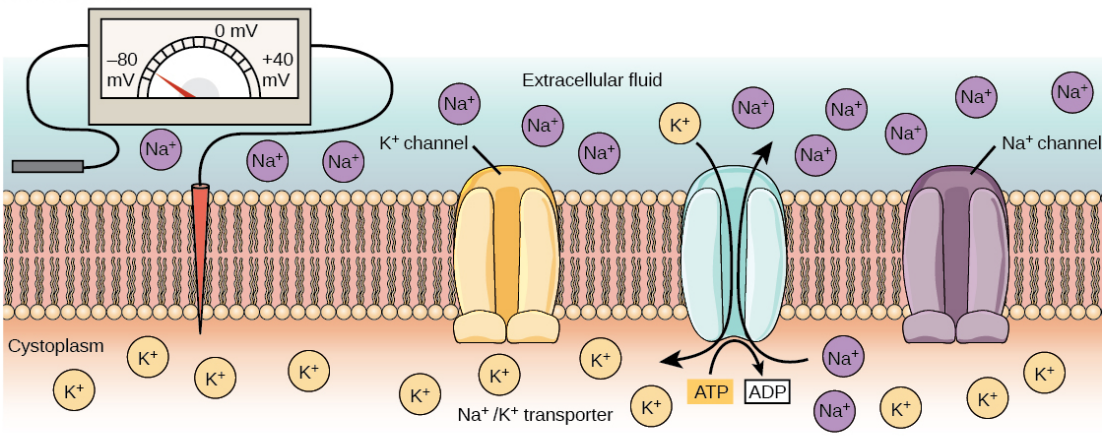
A neuron at rest is negatively charged: the inside of a cell is approximately 70 millivolts more negative than the outside (−70 mV, note that this number varies by neuron type and by species). This voltage is called the resting membrane potential; it is caused by differences in the concentrations of ions inside and outside the cell. If the membrane were equally permeable to all ions, each type of ion would flow across the membrane and the system would reach equilibrium. Because ions cannot simply cross the membrane at will, there are different concentrations of several ions inside and outside the cell, as shown in [link]. The difference in the number of positively charged potassium ions (K^+) inside and outside the cell dominates the resting membrane potential ([link]). When the membrane is at rest, K^+ ions accumulate inside the cell due to a net movement with the concentration gradient. The negative resting membrane potential is created and maintained by increasing the concentration of cations outside the cell (in the extracellular fluid) relative to inside the cell (in the cytoplasm). The negative charge within the cell is created by the cell membrane being more permeable to potassium ion movement than sodium ion movement. In neurons, potassium ions are maintained at high concentrations within the cell while sodium ions are maintained at high concentrations outside of the cell. The cell possesses potassium and sodium leakage channels that allow the two cations to diffuse down their concentration gradient. However, the neurons have far more potassium leakage channels than sodium leakage channels. Therefore, potassium diffuses out of the cell at a much faster rate than sodium leaks in. Because more cations are leaving the cell than are entering, this causes the interior of the cell to be negatively charged relative to the outside of the cell. The actions of the sodium potassium pump help to maintain the resting potential, once established. Recall that sodium potassium pumps brings two K^+ ions into the cell while removing three Na^+ ions per ATP consumed. As more cations are expelled from the cell than taken in, the inside of the cell remains negatively charged relative to the

extracellular fluid. It should be noted that chloride ions (Cl^-) tend to accumulate outside of the cell because they are repelled by negatively-charged proteins within the cytoplasm.

Ion Concentration Inside and Outside Neurons			
Ion	**Extracellular concentration (mM)**	**Intracellular concentration (mM)**	**Ratio outside/inside**
Na^+	145	12	12
K+	4	155	0.026
Cl^-	120	4	30
Organic anions (A^-)	—	100	

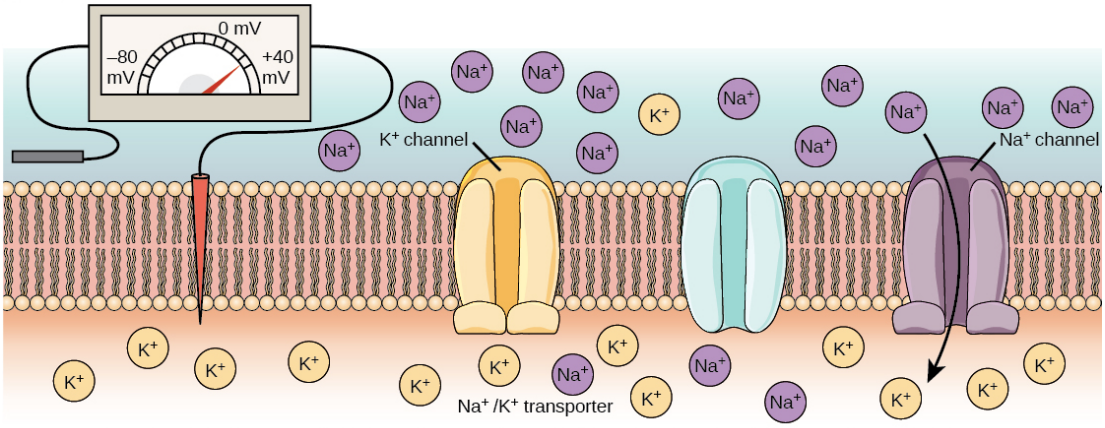
The resting membrane potential is a result of different concentrations inside and outside the cell.

(a) Resting potential



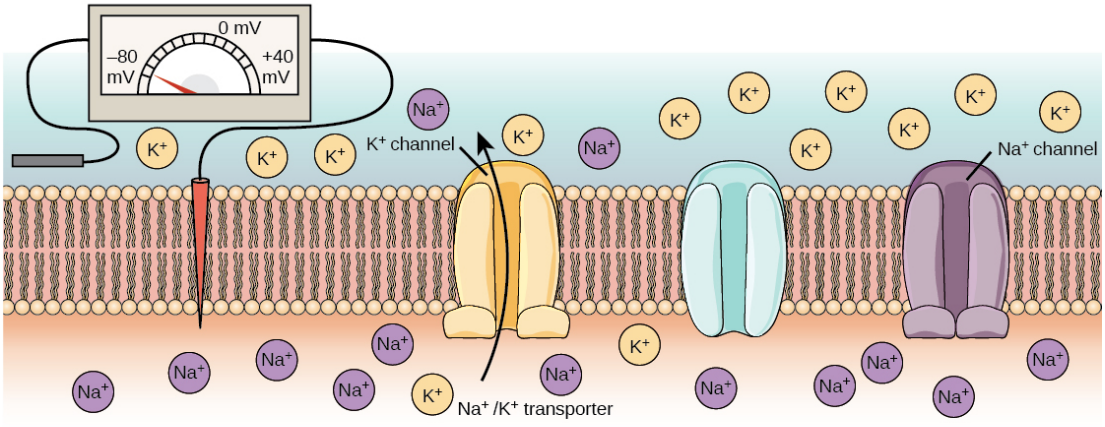
At the resting potential, all voltage-gated Na^+ channels and most voltage-gated K^+ channels are closed. The Na^+/K^+ transporter pumps K^+ ions into the cell and Na^+ ions out.

(b) Depolarization



In response to a depolarization, some Na^+ channels open, allowing Na^+ ions to enter the cell. The membrane starts to depolarize (the charge across the membrane lessens). If the threshold of excitation is reached, all the Na^+ channels open.

(c) Hyperpolarization



At the peak action potential, Na^+ channels close while K^+ channels open. K^+ leaves the cell, and the membrane eventually becomes hyperpolarized.

The (a) resting membrane potential is a result of different concentrations of Na^+ and K^+ ions inside and outside the cell.

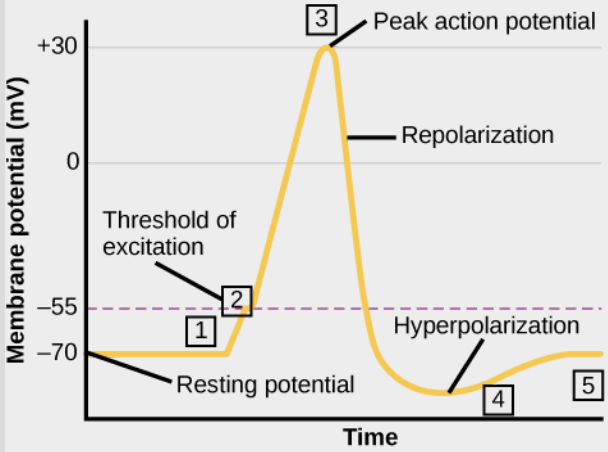
A nerve impulse causes Na^+ to enter the cell, resulting in (b) depolarization. At the peak action potential, K^+ channels open and the cell becomes (c) hyperpolarized.

Action Potential

A neuron can receive input from other neurons and, if this input is strong enough, send the signal to downstream neurons. Transmission of a signal between neurons is generally carried by a chemical called a neurotransmitter. Transmission of a signal within a neuron (from dendrite to axon terminal) is carried by a brief reversal of the resting membrane potential called an **action potential**. When neurotransmitter molecules bind to receptors located on a neuron's dendrites, ion channels open. At excitatory synapses, this opening allows positive ions to enter the neuron and results in **depolarization** of the membrane—a decrease in the difference in voltage between the inside and outside of the neuron. A stimulus from a sensory cell or another neuron depolarizes the target neuron to its threshold potential (-55 mV). Na^+ channels in the axon hillock open, allowing positive ions to enter the cell ([link] and [link]). Once the sodium channels open, the neuron completely depolarizes to a membrane potential of about +40 mV. Action potentials are considered an "all-or nothing" event, in that, once the threshold potential is reached, the neuron always completely depolarizes. Once depolarization is complete, the cell must now "reset" its membrane voltage back to the resting potential. To accomplish this, the Na^+ channels close and cannot be opened. This begins the neuron's **refractory period**, in which it cannot produce another action potential because its sodium channels will not open. At the same time, voltage-gated K^+ channels open, allowing K^+ to leave the cell. As K^+ ions leave the cell, the membrane potential once again becomes negative. The diffusion of K^+ out of the cell actually **hyperpolarizes** the cell, in that the membrane potential becomes more negative than the cell's normal resting potential. At this point, the sodium channels will return to their resting state, meaning they are ready to open again if the membrane potential again exceeds the

threshold potential. Eventually the extra K^+ ions diffuse out of the cell through the potassium leakage channels, bringing the cell from its hyperpolarized state, back to its resting membrane potential.

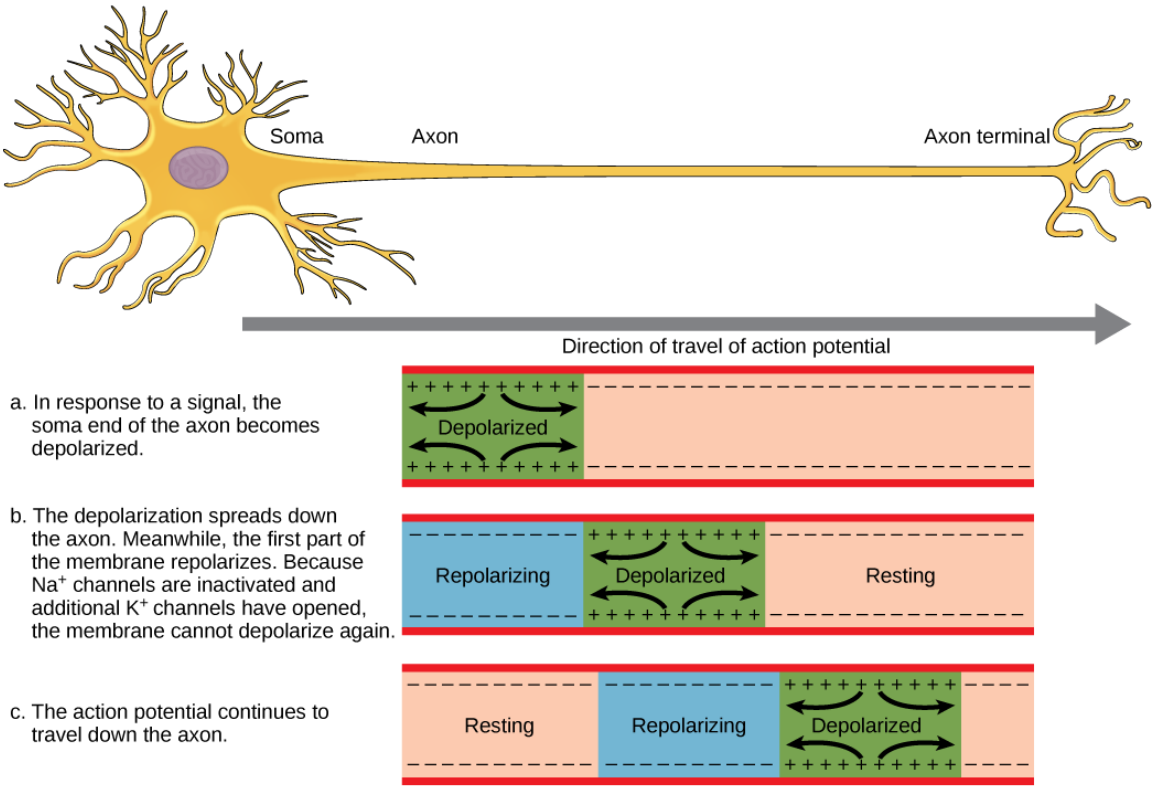
Note:
Art Connection



The formation of an action potential can be divided into five steps: (1) A stimulus from a sensory cell or another neuron causes the target cell to depolarize toward the threshold potential. (2) If the threshold of excitation is reached, all Na^+ channels open and the membrane depolarizes. (3) At the peak action potential, K^+ channels open and K^+ begins to leave the cell. At the same time, Na^+ channels close. (4) The membrane becomes hyperpolarized as K^+ ions continue to leave the cell. The hyperpolarized membrane is in

a refractory period and cannot fire. (5) The K^+ channels close and the Na^+/K^+ transporter restores the resting potential.

Potassium channel blockers, such as amiodarone and procainamide, which are used to treat abnormal electrical activity in the heart, called cardiac dysrhythmia, impede the movement of K^+ through voltage-gated K^+ channels. Which part of the action potential would you expect potassium channels to affect?



The action potential is conducted down the axon as the axon membrane depolarizes, then repolarizes.

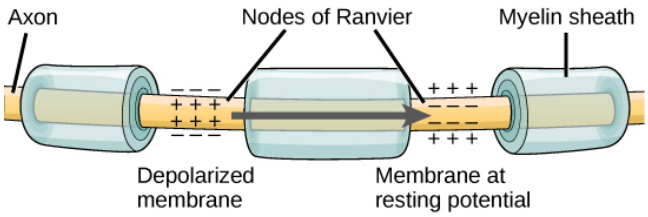
Note:
Link to Learning



This video presents an overview of action potential.

Myelin and the Propagation of the Action Potential

For an action potential to communicate information to another neuron, it must travel along the axon and reach the axon terminals where it can initiate neurotransmitter release. The speed of conduction of an action potential along an axon is influenced by both the diameter of the axon and the axon's resistance to current leak. Myelin acts as an insulator that prevents current from leaving the axon; this increases the speed of action potential conduction. In demyelinating diseases like multiple sclerosis, action potential conduction slows because current leaks from previously insulated axon areas. The nodes of Ranvier, illustrated in [link] are gaps in the myelin sheath along the axon. These unmyelinated spaces are about one micrometer long and contain voltage gated Na^+ and K^+ channels. Flow of ions through these channels, particularly the Na^+ channels, regenerates the action potential over and over again along the axon. This 'jumping' of the action potential from one node to the next is called **saltatory conduction**. If nodes of Ranvier were not present along an axon, the action potential would propagate very slowly since Na^+ and K^+ channels would have to continuously regenerate action potentials at every point along the axon instead of at specific points. Nodes of Ranvier also save energy for the neuron since the channels only need to be present at the nodes and not along the entire axon.



Nodes of Ranvier are gaps in myelin coverage along axons. Nodes contain voltage-gated K^+ and Na^+ channels. Action potentials travel down the axon by jumping from one node to the next.

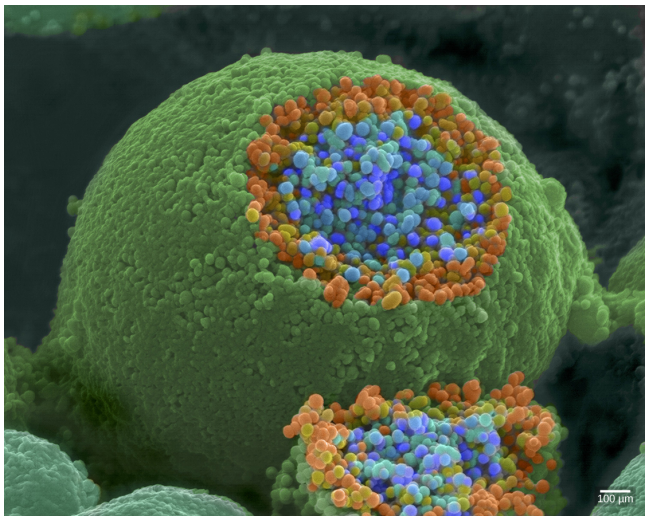
Synaptic Transmission

The synapse or “gap” is the place where information is transmitted from one neuron to another. Synapses usually form between axon terminals and dendritic spines, but this is not universally true. There are also axon-to-axon, dendrite-to-dendrite, and axon-to-cell body synapses. The neuron transmitting the signal is called the presynaptic neuron, and the neuron receiving the signal is called the postsynaptic neuron. Note that these designations are relative to a particular synapse—most neurons are both presynaptic and postsynaptic. There are two types of synapses: chemical and electrical.

Chemical Synapse

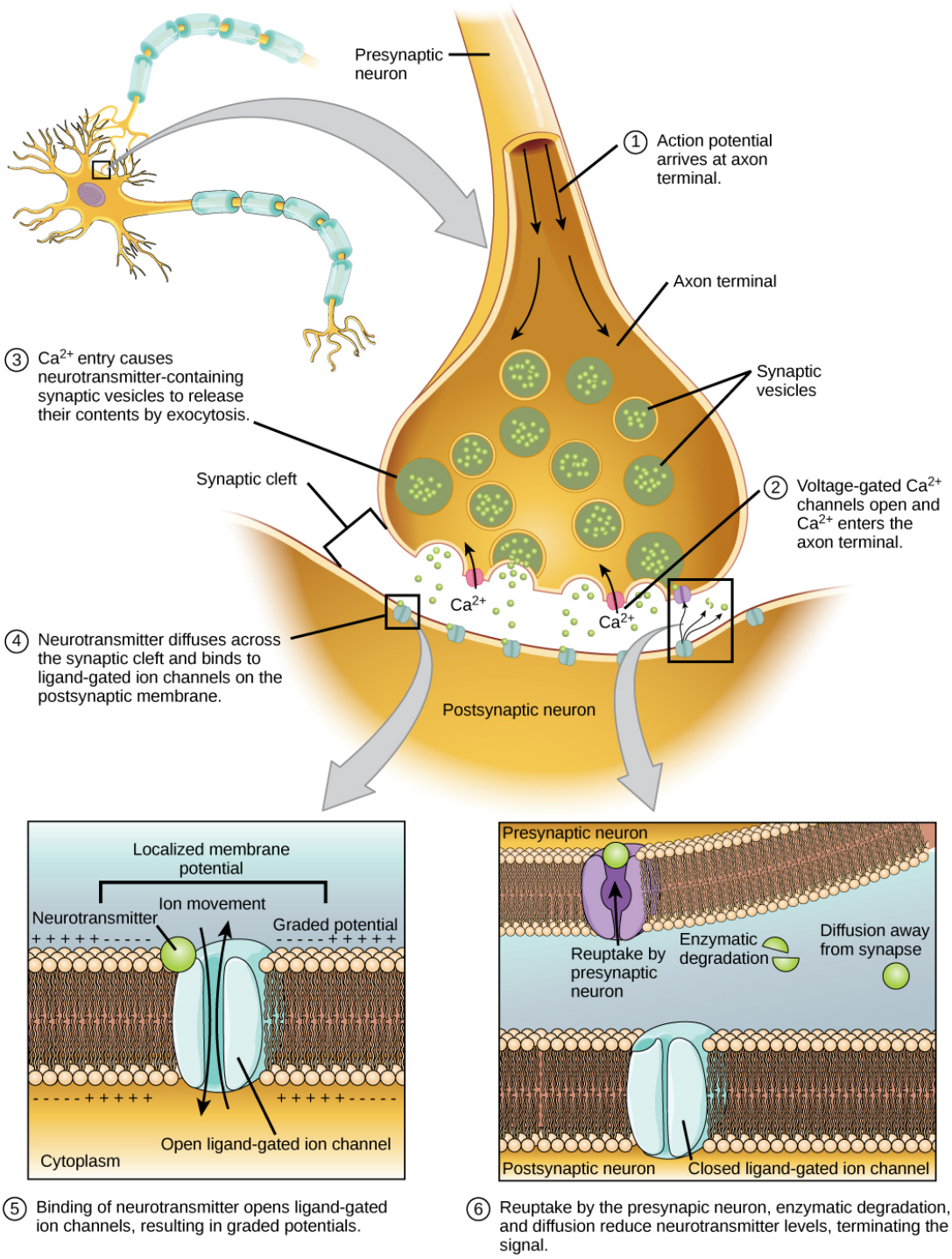
When an action potential reaches the axon terminal it depolarizes the membrane and opens voltage-gated Na^+ channels. Na^+ ions enter the cell, further depolarizing the presynaptic membrane. This depolarization causes voltage-gated Ca^{2+} channels to open. Calcium ions entering the cell initiate a signaling cascade that causes small membrane-bound vesicles, called

synaptic vesicles, containing neurotransmitter molecules to fuse with the presynaptic membrane. Synaptic vesicles are shown in [link], which is an image from a scanning electron microscope.



This pseudocolored image taken with a scanning electron microscope shows an axon terminal that was broken open to reveal synaptic vesicles (blue and orange) inside the neuron. (credit: modification of work by Tina Carvalho, NIH-NIGMS; scale-bar data from Matt Russell)

Fusion of a vesicle with the presynaptic membrane causes neurotransmitter to be released into the **synaptic cleft**, the extracellular space between the presynaptic and postsynaptic membranes, as illustrated in [link]. The neurotransmitter diffuses across the synaptic cleft and binds to receptor proteins on the postsynaptic membrane.



Communication at chemical synapses requires release of neurotransmitters. When the presynaptic membrane is depolarized, voltage-gated Ca^{2+} channels open and allow Ca^{2+} to enter the cell. The calcium entry causes synaptic vesicles to fuse

with the membrane and release neurotransmitter molecules into the synaptic cleft. The neurotransmitter diffuses across the synaptic cleft and binds to ligand-gated ion channels in the postsynaptic membrane, resulting in a localized depolarization or hyperpolarization of the postsynaptic neuron.

The binding of a specific neurotransmitter causes particular ion channels, in this case ligand-gated channels, on the postsynaptic membrane to open. Neurotransmitters can either have excitatory or inhibitory effects on the postsynaptic membrane, as detailed in [link]. For example, when acetylcholine is released at the synapse between a nerve and muscle (called the neuromuscular junction) by a presynaptic neuron, it causes postsynaptic Na^+ channels to open. Na^+ enters the postsynaptic cell and causes the postsynaptic membrane to depolarize. This depolarization is called an **excitatory postsynaptic potential (EPSP)** and makes the postsynaptic neuron more likely to fire an action potential. Release of neurotransmitter at inhibitory synapses causes **inhibitory postsynaptic potentials (IPSPs)**, a hyperpolarization of the presynaptic membrane. For example, when the neurotransmitter GABA (gamma-aminobutyric acid) is released from a presynaptic neuron, it binds to and opens Cl^- channels. Cl^- ions enter the cell and hyperpolarizes the membrane, making the neuron less likely to fire an action potential.

Once neurotransmission has occurred, the neurotransmitter must be removed from the synaptic cleft so the postsynaptic membrane can “reset” and be ready to receive another signal. This can be accomplished in three ways: the neurotransmitter can diffuse away from the synaptic cleft, it can be degraded by enzymes in the synaptic cleft, or it can be recycled (sometimes called reuptake) by the presynaptic neuron. Several drugs act at this step of neurotransmission. For example, some drugs that are given to Alzheimer’s patients work by inhibiting acetylcholinesterase, the enzyme that degrades acetylcholine. This inhibition of the enzyme essentially increases neurotransmission at synapses that release acetylcholine. Once released, the acetylcholine stays in the cleft and can continually bind and unbind to postsynaptic receptors.

Neurotransmitter Function and Location		
Neurotransmitter	**Example**	**Location**
Acetylcholine	—	CNS and/or PNS
Biogenic amine	Dopamine, serotonin, norepinephrine	CNS and/or PNS
Amino acid	Glycine, glutamate, aspartate, gamma aminobutyric acid	CNS
Neuropeptide	Substance P, endorphins	CNS and/or PNS

Electrical Synapse

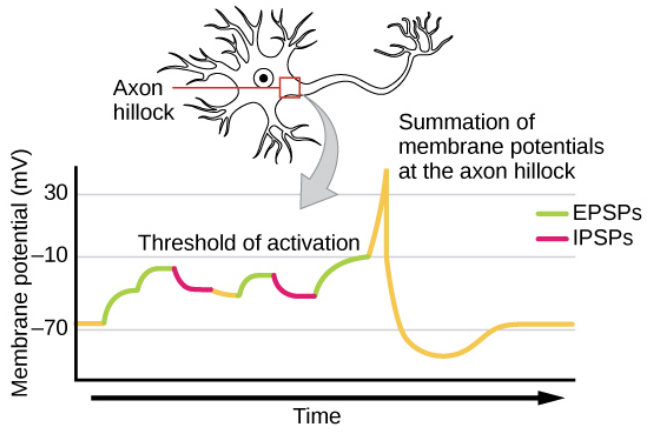
While electrical synapses are fewer in number than chemical synapses, they are found in all nervous systems and play important and unique roles. The mode of neurotransmission in electrical synapses is quite different from that in chemical synapses. In an electrical synapse, the presynaptic and postsynaptic membranes are very close together and are actually physically connected by channel proteins forming gap junctions. Gap junctions allow current to pass directly from one cell to the next. In addition to the ions that carry this current, other molecules, such as ATP, can diffuse through the large gap junction pores.

There are key differences between chemical and electrical synapses. Because chemical synapses depend on the release of neurotransmitter molecules from synaptic vesicles to pass on their signal, there is an approximately one millisecond delay between when the axon potential

reaches the presynaptic terminal and when the neurotransmitter leads to opening of postsynaptic ion channels. Additionally, this signaling is unidirectional. Signaling in electrical synapses, in contrast, is virtually instantaneous (which is important for synapses involved in key reflexes), and some electrical synapses are bidirectional. Electrical synapses are also more reliable as they are less likely to be blocked, and they are important for synchronizing the electrical activity of a group of neurons. For example, electrical synapses in the thalamus are thought to regulate slow-wave sleep, and disruption of these synapses can cause seizures.

Signal Summation

Sometimes a single EPSP is strong enough to induce an action potential in the postsynaptic neuron, but often multiple presynaptic inputs must create EPSPs around the same time for the postsynaptic neuron to be sufficiently depolarized to fire an action potential. This process is called **summation** and occurs at the axon hillock, as illustrated in [link]. Additionally, one neuron often has inputs from many presynaptic neurons—some excitatory and some inhibitory—so IPSPs can cancel out EPSPs and vice versa. It is the net change in postsynaptic membrane voltage that determines whether the postsynaptic cell has reached its threshold of excitation needed to fire an action potential. Together, synaptic summation and the threshold for excitation act as a filter so that random "noise" in the system is not transmitted as important information.

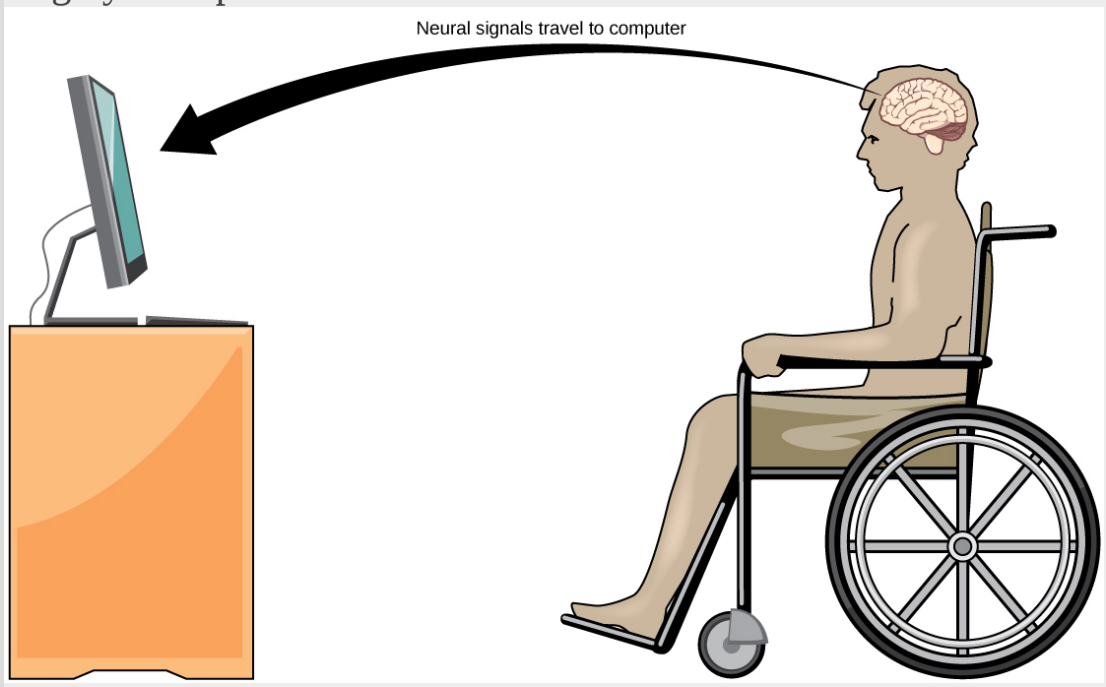


A single neuron can receive both excitatory and inhibitory inputs from multiple neurons, resulting in local membrane depolarization (EPSP input) and hyperpolarization (IPSP input). All these inputs are added together at the axon hillock. If the EPSPs are strong enough to overcome the IPSPs and reach the threshold of excitation, the neuron will fire.

Note:
Everyday Connection
Brain-computer interface
Amyotrophic lateral sclerosis (ALS, also called Lou Gehrig's Disease) is a neurological disease characterized by the degeneration of the motor neurons that control voluntary movements. The disease begins with muscle weakening and lack of coordination and eventually destroys the neurons that control speech, breathing, and swallowing; in the end, the disease can lead to paralysis. At that point, patients require assistance from machines to be able to breathe and to communicate. Several special technologies have been developed to allow "locked-in" patients to communicate with the rest of the world. One technology, for example, allows patients to type out sentences by twitching their cheek. These sentences can then be read aloud by a computer.
A relatively new line of research for helping paralyzed patients, including those with ALS, to communicate and retain a degree of self-sufficiency is called brain-computer interface (BCI) technology and is illustrated in [link]. This technology sounds like something out of science fiction: it allows paralyzed patients to control a computer using only their thoughts. There are several forms of BCI. Some forms use EEG recordings from electrodes taped onto the skull. These recordings contain information from

large populations of neurons that can be decoded by a computer. Other forms of BCI require the implantation of an array of electrodes smaller than a postage stamp in the arm and hand area of the motor cortex. This form of BCI, while more invasive, is very powerful as each electrode can record actual action potentials from one or more neurons. These signals are then sent to a computer, which has been trained to decode the signal and feed it to a tool—such as a cursor on a computer screen. This means that a patient with ALS can use e-mail, read the Internet, and communicate with others by thinking of moving his or her hand or arm (even though the paralyzed patient cannot make that bodily movement). Recent advances have allowed a paralyzed locked-in patient who suffered a stroke 15 years ago to control a robotic arm and even to feed herself coffee using BCI technology.

Despite the amazing advancements in BCI technology, it also has limitations. The technology can require many hours of training and long periods of intense concentration for the patient; it can also require brain surgery to implant the devices.



With brain-computer interface technology, neural signals from a paralyzed patient are collected, decoded, and then fed to a tool, such as a computer, a wheelchair, or a robotic arm.

Note:
Link to Learning



Watch this video in which a paralyzed woman use a brain-controlled robotic arm to bring a drink to her mouth, among other images of brain-computer interface technology in action.
https://www.openstaxcollege.org/l/paralyzation

Synaptic Plasticity

Synapses are not static structures. They can be weakened or strengthened. They can be broken, and new synapses can be made. Synaptic plasticity allows for these changes, which are all needed for a functioning nervous system. In fact, synaptic plasticity is the basis of learning and memory. Two processes in particular, long-term potentiation (LTP) and long-term depression (LTD) are important forms of synaptic plasticity that occur in synapses in the hippocampus, a brain region that is involved in storing memories.

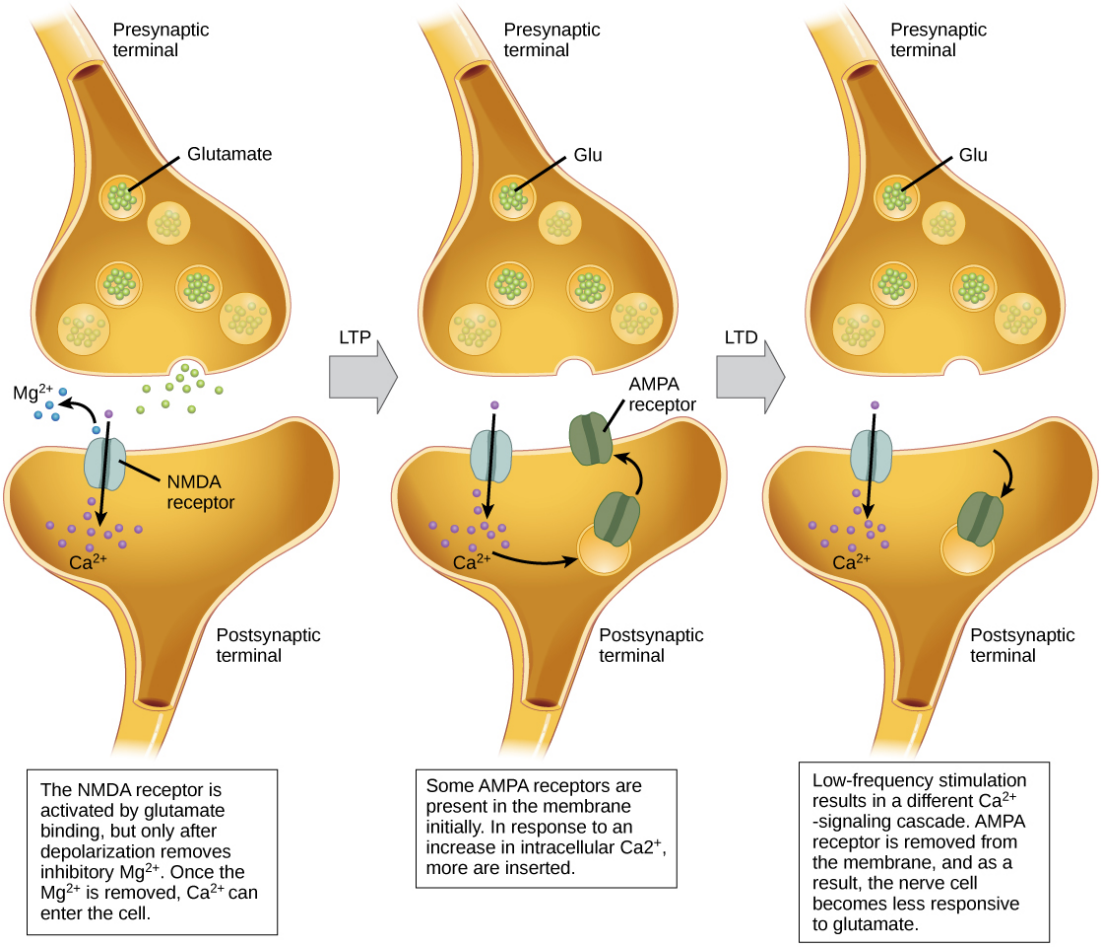
Long-term Potentiation (LTP)

Long-term potentiation (LTP) is a persistent strengthening of a synaptic connection. LTP is based on the Hebbian principle: cells that fire together wire together. There are various mechanisms, none fully understood, behind the synaptic strengthening seen with LTP. One known mechanism involves

a type of postsynaptic glutamate receptor, called NMDA (N-Methyl-D-aspartate) receptors, shown in [link]. These receptors are normally blocked by magnesium ions; however, when the postsynaptic neuron is depolarized by multiple presynaptic inputs in quick succession (either from one neuron or multiple neurons), the magnesium ions are forced out allowing Ca ions to pass into the postsynaptic cell. Next, Ca^{2+} ions entering the cell initiate a signaling cascade that causes a different type of glutamate receptor, called AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid) receptors, to be inserted into the postsynaptic membrane, since activated AMPA receptors allow positive ions to enter the cell. So, the next time glutamate is released from the presynaptic membrane, it will have a larger excitatory effect (EPSP) on the postsynaptic cell because the binding of glutamate to these AMPA receptors will allow more positive ions into the cell. The insertion of additional AMPA receptors strengthens the synapse and means that the postsynaptic neuron is more likely to fire in response to presynaptic neurotransmitter release. Some drugs of abuse co-opt the LTP pathway, and this synaptic strengthening can lead to addiction.

Long-term Depression (LTD)

Long-term depression (LTD) is essentially the reverse of LTP: it is a long-term weakening of a synaptic connection. One mechanism known to cause LTD also involves AMPA receptors. In this situation, calcium that enters through NMDA receptors initiates a different signaling cascade, which results in the removal of AMPA receptors from the postsynaptic membrane, as illustrated in [link]. The decrease in AMPA receptors in the membrane makes the postsynaptic neuron less responsive to glutamate released from the presynaptic neuron. While it may seem counterintuitive, LTD may be just as important for learning and memory as LTP. The weakening and pruning of unused synapses allows for unimportant connections to be lost and makes the synapses that have undergone LTP that much stronger by comparison.



Calcium entry through postsynaptic NMDA receptors can initiate two different forms of synaptic plasticity: long-term potentiation (LTP) and long-term depression (LTD). LTP arises when a single synapse is repeatedly stimulated. This stimulation causes a calcium- and CaMKII-dependent cellular cascade, which results in the insertion of more AMPA receptors into the postsynaptic membrane. The next time glutamate is released from the presynaptic cell, it will bind to both NMDA and the newly inserted AMPA receptors, thus depolarizing the membrane more efficiently. LTD occurs when few glutamate molecules bind to NMDA receptors at a synapse (due to a low firing rate of the presynaptic neuron). The calcium that does flow through NMDA receptors initiates a different calcineurin and protein phosphatase 1-dependent cascade, which results in the endocytosis of AMPA receptors.

This makes the postsynaptic neuron less responsive to glutamate released from the presynaptic neuron.

Section Summary

Neurons have charged membranes because there are different concentrations of ions inside and outside of the cell. Voltage-gated ion channels control the movement of ions into and out of a neuron. When a neuronal membrane is depolarized to at least the threshold of excitation, an action potential is fired. The action potential is then propagated along a myelinated axon to the axon terminals. In a chemical synapse, the action potential causes release of neurotransmitter molecules into the synaptic cleft. Through binding to postsynaptic receptors, the neurotransmitter can cause excitatory or inhibitory postsynaptic potentials by depolarizing or hyperpolarizing, respectively, the postsynaptic membrane. In electrical synapses, the action potential is directly communicated to the postsynaptic cell through gap junctions—large channel proteins that connect the pre-and postsynaptic membranes. Synapses are not static structures and can be strengthened and weakened. Two mechanisms of synaptic plasticity are long-term potentiation and long-term depression.

Art Connections

Exercise:

Problem:

[link] Potassium channel blockers, such as amiodarone and procainamide, which are used to treat abnormal electrical activity in the heart, called cardiac dysrhythmia, impede the movement of K+ through voltage-gated K+ channels. Which part of the action potential would you expect potassium channels to affect?

Solution:

[link] Potassium channel blockers slow the repolarization phase, but have no effect on depolarization.

Review Questions

Exercise:

Problem:

For a neuron to fire an action potential, its membrane must reach _____.

a. hyperpolarization
b. the threshold of excitation
c. the refractory period
d. inhibitory postsynaptic potential

Solution:

B

Exercise:

Problem:

After an action potential, the opening of additional voltage-gated _____ channels and the inactivation of sodium channels, cause the membrane to return to its resting membrane potential.

a. sodium
b. potassium
c. calcium
d. chloride

Solution:

B

Exercise:

Problem:

What is the term for protein channels that connect two neurons at an electrical synapse?

a. synaptic vesicles
b. voltage-gated ion channels
c. gap junction protein
d. sodium-potassium exchange pumps

Solution:

C

Free Response

Exercise:

Problem:

How does myelin aid propagation of an action potential along an axon? How do the nodes of Ranvier help this process?

Solution:

Myelin prevents the leak of current from the axon. Nodes of Ranvier allow the action potential to be regenerated at specific points along the axon. They also save energy for the cell since voltage-gated ion channels and sodium-potassium transporters are not needed along myelinated portions of the axon.

Exercise:

Problem: What are the main steps in chemical neurotransmission?

Solution:

An action potential travels along an axon until it depolarizes the membrane at an axon terminal. Depolarization of the membrane causes voltage-gated Ca^{2+} channels to open and Ca^{2+} to enter the cell. The intracellular calcium influx causes synaptic vesicles containing neurotransmitter to fuse with the presynaptic membrane. The neurotransmitter diffuses across the synaptic cleft and binds to receptors on the postsynaptic membrane. Depending on the specific neurotransmitter and postsynaptic receptor, this action can cause positive (excitatory postsynaptic potential) or negative (inhibitory postsynaptic potential) ions to enter the cell.

Glossary

action potential
: self-propagating momentary change in the electrical potential of a neuron (or muscle) membrane

depolarization
: change in the membrane potential to a less negative value

excitatory postsynaptic potential (EPSP)
: depolarization of a postsynaptic membrane caused by neurotransmitter molecules released from a presynaptic cell

hyperpolarization
: change in the membrane potential to a more negative value

inhibitory postsynaptic potential (IPSP)
: hyperpolarization of a postsynaptic membrane caused by neurotransmitter molecules released from a presynaptic cell

long-term depression (LTD)
: prolonged decrease in synaptic coupling between a pre- and postsynaptic cell

long-term potentiation (LTP)
: prolonged increase in synaptic coupling between a pre-and postsynaptic cell

membrane potential
: difference in electrical potential between the inside and outside of a cell

refractory period
: period after an action potential when it is more difficult or impossible for an action potential to be fired; caused by inactivation of sodium channels and activation of additional potassium channels of the membrane

saltatory conduction
: “jumping” of an action potential along an axon from one node of Ranvier to the next

summation
: process of multiple presynaptic inputs creating EPSPs around the same time for the postsynaptic neuron to be sufficiently depolarized to fire an action potential

synaptic cleft
: space between the presynaptic and postsynaptic membranes

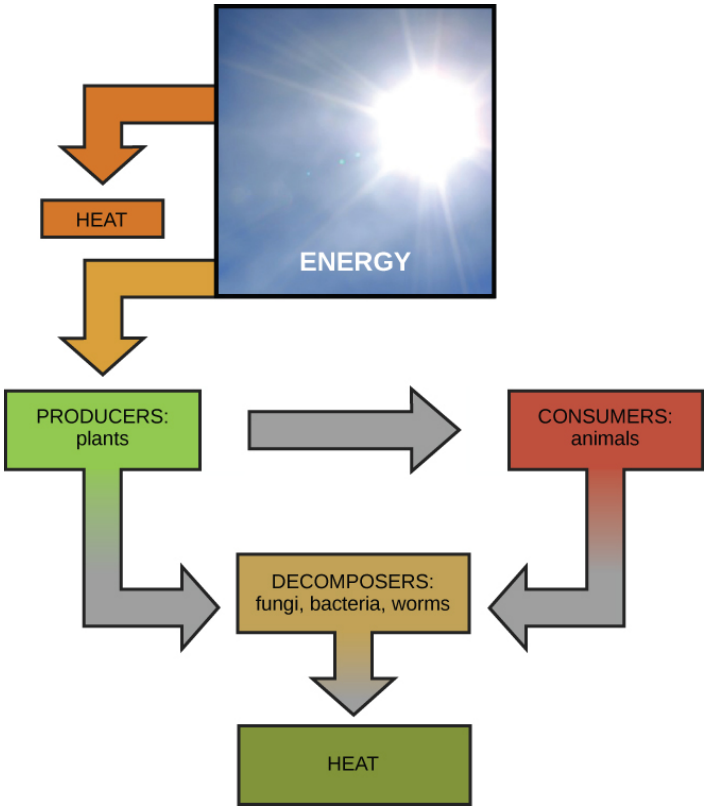
synaptic vesicle
: spherical structure that contains a neurotransmitter

threshold of excitation
: level of depolarization needed for an action potential to fire

Energy and Metabolism
By the end of this section, you will be able to:

- Explain what metabolic pathways are
- State the first and second laws of thermodynamics
- Explain the difference between kinetic and potential energy
- Describe endergonic and exergonic reactions
- Discuss how enzymes function as molecular catalysts

Scientists use the term **bioenergetics** to describe the concept of energy flow ([link]) through living systems, such as cells. Cellular processes such as the building and breaking down of complex molecules occur through stepwise chemical reactions. Some of these chemical reactions are spontaneous and release energy, whereas others require energy to proceed. Just as living things must continually consume food to replenish their energy supplies, cells must continually produce more energy to replenish that used by the many energy-requiring chemical reactions that constantly take place. Together, all of the chemical reactions that take place inside cells, including those that consume or generate energy, are referred to as the cell's **metabolism**.



Ultimately, most life forms get their energy from the sun. Plants use photosynthesis to capture sunlight, and herbivores eat the plants to obtain energy. Carnivores eat the herbivores, and eventual decomposition of plant and animal material contributes to the nutrient pool.

Metabolic Pathways

Consider the metabolism of sugar. This is a classic example of one of the many cellular processes that use and produce energy. Living things consume sugars as a major energy source, because sugar molecules have a great deal of energy stored within their bonds. For the most part, photosynthesizing organisms like plants produce these sugars. During photosynthesis, plants use energy (originally from sunlight) to convert carbon dioxide gas (CO_2) into sugar molecules (like glucose: $C_6H_{12}O_6$). They consume carbon dioxide and produce oxygen as a waste product. This reaction is summarized as:

Equation:

$$6CO_2 + 6H_2O \text{ --> } C_6H_{12}O_6 + 6O_2$$

Because this process involves synthesizing an energy-storing molecule, it requires energy input to proceed. During the light reactions of photosynthesis, energy is provided by a molecule called adenosine triphosphate (ATP), which is the primary energy currency of all cells. Just as the dollar is used as currency to buy goods, cells use molecules of ATP as energy currency to perform immediate work. In contrast, energy-storage molecules such as glucose are consumed only to be broken down to use their energy. The reaction that harvests the energy of a sugar molecule in cells requiring oxygen to survive can be summarized by the reverse reaction

to photosynthesis. In this reaction, oxygen is consumed and carbon dioxide is released as a waste product. The reaction is summarized as:
Equation:

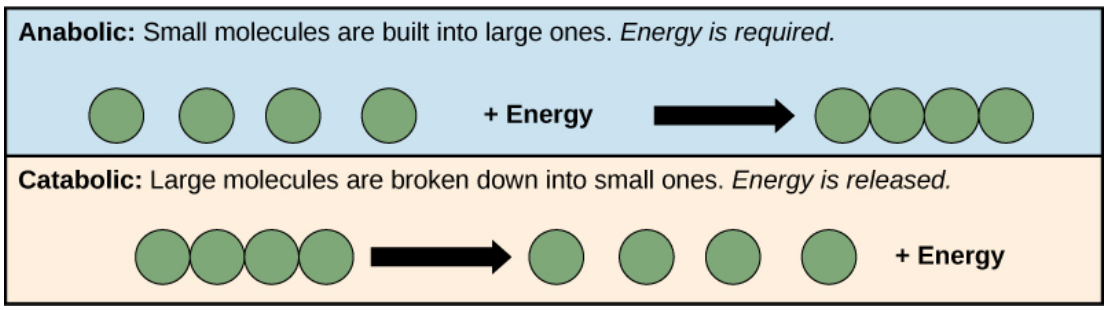
$$C_6H_{12}O_6 + 6O_2 \text{ --> } 6H_2O + 6CO_2$$

Both of these reactions involve many steps.

The processes of making and breaking down sugar molecules illustrate two examples of metabolic pathways. A metabolic pathway is a series of chemical reactions that takes a starting molecule and modifies it, step-by-step, through a series of metabolic intermediates, eventually yielding a final product. In the example of sugar metabolism, the first metabolic pathway synthesized sugar from smaller molecules, and the other pathway broke sugar down into smaller molecules. These two opposite processes—the first requiring energy and the second producing energy—are referred to as **anabolic** pathways (building polymers) and **catabolic** pathways (breaking down polymers into their monomers), respectively. Consequently, metabolism is composed of synthesis (anabolism) and degradation (catabolism) ([link]).

It is important to know that the chemical reactions of metabolic pathways do not take place on their own. Each reaction step is facilitated, or catalyzed, by a protein called an enzyme. Enzymes are important for catalyzing all types of biological reactions—those that require energy as well as those that release energy.

Metabolic pathways



Catabolic pathways are those that generate energy by breaking down larger molecules. Anabolic pathways are those that require energy to synthesize larger molecules. Both types of pathways are required for maintaining the cell's energy balance.

Energy

Thermodynamics refers to the study of energy and energy transfer involving physical matter. The matter relevant to a particular case of energy transfer is called a system, and everything outside of that matter is called the surroundings. For instance, when heating a pot of water on the stove, the system includes the stove, the pot, and the water. Energy is transferred within the system (between the stove, pot, and water). There are two types of systems: open and closed. In an open system, energy can be exchanged with its surroundings. The stovetop system is open because heat can be lost to the air. A closed system cannot exchange energy with its surroundings.

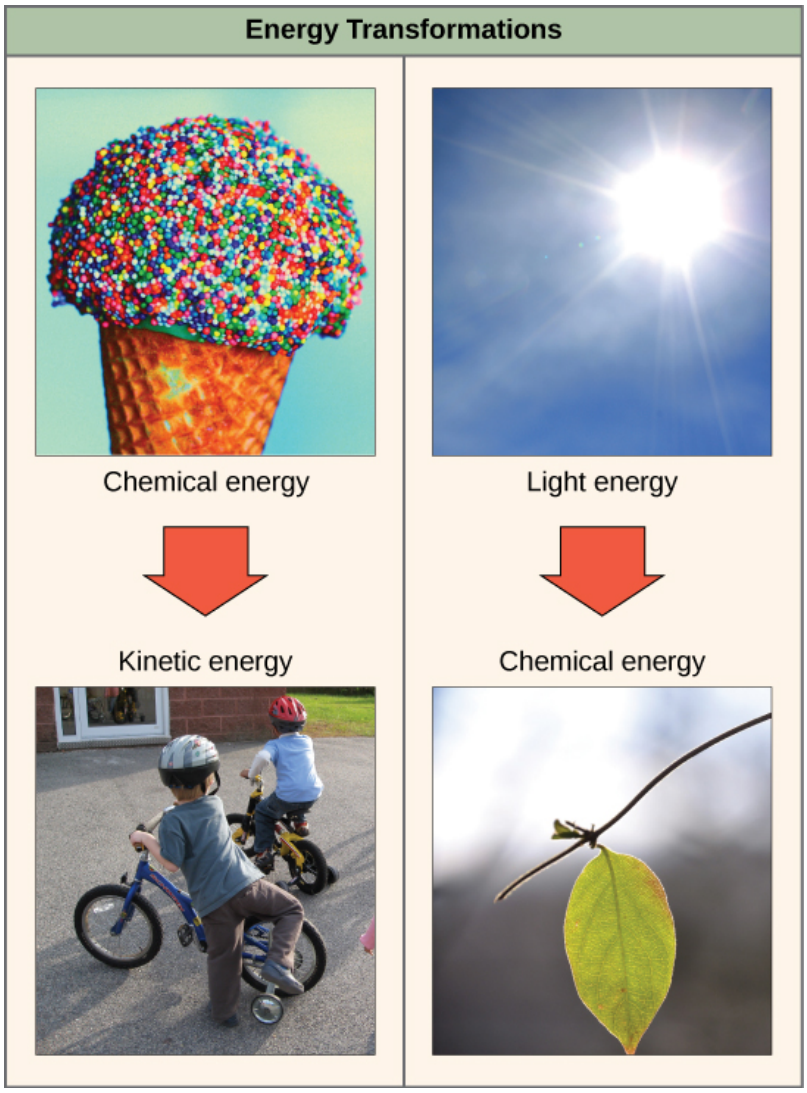
Biological organisms are open systems. Energy is exchanged between them and their surroundings as they use energy from the sun to perform photosynthesis or consume energy-storing molecules and release energy to the environment by doing work and releasing heat. Like all things in the physical world, energy is subject to physical laws. The laws of thermodynamics govern the transfer of energy in and among all systems in the universe.

In general, energy is defined as the ability to do work, or to create some kind of change. Energy exists in different forms. For example, electrical energy, light energy, and heat energy are all different types of energy. To appreciate the way energy flows into and out of biological systems, it is important to understand two of the physical laws that govern energy.

Thermodynamics

The first law of thermodynamics states that the total amount of energy in the universe is constant and conserved. In other words, there has always been, and always will be, exactly the same amount of energy in the universe. Energy exists in many different forms. According to the first law of thermodynamics, energy may be transferred from place to place or transformed into different forms, but it cannot be created or destroyed. The transfers and transformations of energy take place around us all the time. Light bulbs transform electrical energy into light and heat energy. Gas stoves transform chemical energy from natural gas into heat energy. Plants perform one of the most biologically useful energy transformations on earth: that of converting the energy of sunlight to chemical energy stored within organic molecules ([link]). Some examples of energy transformations are shown in [link].

The challenge for all living organisms is to obtain energy from their surroundings in forms that they can transfer or transform into usable energy to do work. Living cells have evolved to meet this challenge. Chemical energy stored within organic molecules such as sugars and fats is transferred and transformed through a series of cellular chemical reactions into energy within molecules of ATP. Energy in ATP molecules is easily accessible to do work. Examples of the types of work that cells need to do include building complex molecules, transporting materials, powering the motion of cilia or flagella, and contracting muscle fibers to create movement.



Shown are some examples of energy transferred and transformed from one system to another and from one form to another. The food we consume provides our cells with the energy required to carry out bodily functions, just as light energy provides plants with the means to create the chemical energy they need. (credit "ice cream": modification of work by D. Sharon Pruitt; credit "kids": modification of work by Max from Providence; credit "leaf": modification of work by Cory Zanker)

A living cell's primary tasks of obtaining, transforming, and using energy to do work may seem simple. However, the second law of thermodynamics explains why these tasks are harder than they appear. All energy transfers and transformations are never completely efficient. In every energy transfer, some amount of energy is lost in a form that is unusable. In most cases, this form is heat energy. Thermodynamically, **heat energy** is defined as the energy transferred from one system to another that is not work. For example, when a light bulb is turned on, some of the energy being converted from electrical energy into light energy is lost as heat energy. Likewise, some energy is lost as heat energy during cellular metabolic reactions.

An important concept in physical systems is that of order and disorder. The more energy that is lost by a system to its surroundings, the less ordered and more random the system is. Scientists refer to the measure of randomness or disorder within a system as entropy. High entropy means high disorder and low energy. Molecules and chemical reactions have varying entropy as well. For example, entropy increases as molecules at a high concentration in one place diffuse and spread out. The second law of thermodynamics says that energy will always be lost as heat in energy transfers or transformations.

Living things are highly ordered, requiring constant energy input to be maintained in a state of low entropy.

Potential and Kinetic Energy

When an object is in motion, there is energy associated with that object. Think of a wrecking ball. Even a slow-moving wrecking ball can do a great deal of damage to other objects. Energy associated with objects in motion is called **kinetic energy** ([link]). A speeding bullet, a walking person, and the rapid movement of molecules in the air (which produces heat) all have kinetic energy.

Now what if that same motionless wrecking ball is lifted two stories above ground with a crane? If the suspended wrecking ball is unmoving, is there

energy associated with it? The answer is yes. The energy that was required to lift the wrecking ball did not disappear, but is now stored in the wrecking ball by virtue of its position and the force of gravity acting on it. This type of energy is called **potential energy** ([link]). If the ball were to fall, the potential energy would be transformed into kinetic energy until all of the potential energy was exhausted when the ball rested on the ground. Wrecking balls also swing like a pendulum; through the swing, there is a constant change of potential energy (highest at the top of the swing) to kinetic energy (highest at the bottom of the swing). Other examples of potential energy include the energy of water held behind a dam or a person about to skydive out of an airplane.



Still water has potential energy; moving water, such as in a waterfall or a rapidly flowing river, has kinetic energy. (credit "dam": modification of work by "Pascal"/Flickr; credit "waterfall": modification of work by Frank Gualtieri)

Potential energy is not only associated with the location of matter, but also with the structure of matter. Even a spring on the ground has potential energy if it is compressed; so does a rubber band that is pulled taut. On a molecular level, the bonds that hold the atoms of molecules together exist in a particular structure that has potential energy. Remember that anabolic

cellular pathways require energy to synthesize complex molecules from simpler ones and catabolic pathways release energy when complex molecules are broken down. The fact that energy can be released by the breakdown of certain chemical bonds implies that those bonds have potential energy. In fact, there is potential energy stored within the bonds of all the food molecules we eat, which is eventually harnessed for use. This is because these bonds can release energy when broken. The type of potential energy that exists within chemical bonds, and is released when those bonds are broken, is called chemical energy. Chemical energy is responsible for providing living cells with energy from food. The release of energy occurs when the molecular bonds within food molecules are broken.

Note:
Concept in Action



Visit the site and select “Pendulum” from the “Work and Energy” menu to see the shifting kinetic and potential energy of a pendulum in motion.

Free and Activation Energy

After learning that chemical reactions release energy when energy-storing bonds are broken, an important next question is the following: How is the energy associated with these chemical reactions quantified and expressed? How can the energy released from one reaction be compared to that of another reaction? A measurement of free energy is used to quantify these energy transfers. Recall that according to the second law of thermodynamics, all energy transfers involve the loss of some amount of energy in an unusable form such as heat. Free energy specifically refers to

the energy associated with a chemical reaction that is available after the losses are accounted for. In other words, free energy is usable energy, or energy that is available to do work.

If energy is released during a chemical reaction, then the change in free energy, signified as ΔG (delta G) will be a negative number. A negative change in free energy also means that the products of the reaction have less free energy than the reactants, because they release some free energy during the reaction. Reactions that have a negative change in free energy and consequently release free energy are called **exergonic reactions**. Think: *ex*ergonic means energy is *ex*iting the system. These reactions are also referred to as spontaneous reactions, and their products have less stored energy than the reactants. An important distinction must be drawn between the term spontaneous and the idea of a chemical reaction occurring immediately. Contrary to the everyday use of the term, a spontaneous reaction is not one that suddenly or quickly occurs. The rusting of iron is an example of a spontaneous reaction that occurs slowly, little by little, over time.

If a chemical reaction absorbs energy rather than releases energy on balance, then the ΔG for that reaction will be a positive value. In this case, the products have more free energy than the reactants. Thus, the products of these reactions can be thought of as energy-storing molecules. These chemical reactions are called **endergonic reactions** and they are non-spontaneous. An endergonic reaction will not take place on its own without the addition of free energy.

Note:
Art Connection



(a)



(b)



(c)



(d)

Shown are some examples of endergonic processes (ones that require energy) and exergonic processes (ones that release energy). (credit a: modification of work by Natalie Maynor; credit b: modification of work by USDA; credit c: modification of work by Cory Zanker; credit d: modification of work by Harry Malsch)

Look at each of the processes shown and decide if it is endergonic or exergonic.

There is another important concept that must be considered regarding endergonic and exergonic reactions. Exergonic reactions require a small amount of energy input to get going, before they can proceed with their

energy-releasing steps. These reactions have a net release of energy, but still require some energy input in the beginning. This small amount of energy input necessary for all chemical reactions to occur is called the **activation energy**.

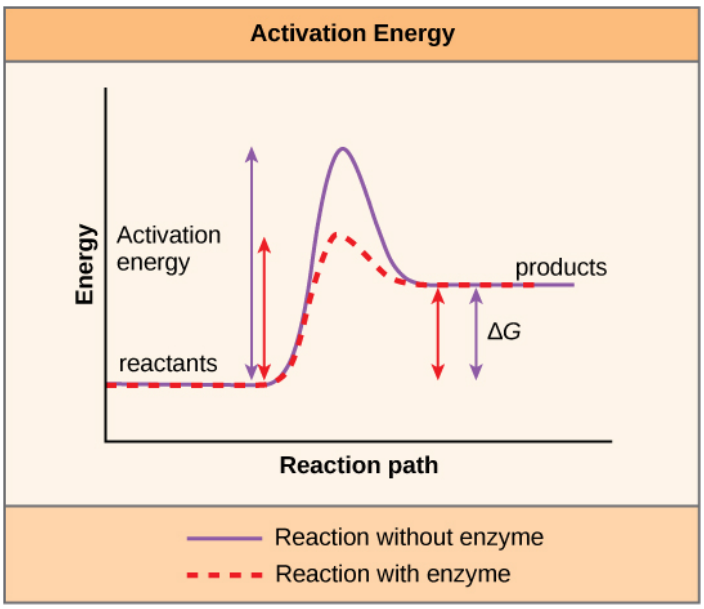
Note:
Concept in Action



Watch an animation of the move from free energy to transition state of the reaction.

Enzymes

A substance that helps a chemical reaction to occur is called a catalyst, and the molecules that catalyze biochemical reactions are called **enzymes**. Most enzymes are proteins and perform the critical task of lowering the activation energies of chemical reactions inside the cell. Most of the reactions critical to a living cell happen too slowly at normal temperatures to be of any use to the cell. Without enzymes to speed up these reactions, life could not persist. Enzymes do this by binding to the reactant molecules and holding them in such a way as to make the chemical bond-breaking and -forming processes take place more easily. It is important to remember that enzymes do not change whether a reaction is exergonic (spontaneous) or endergonic. This is because they do not change the free energy of the reactants or products. They only reduce the activation energy required for the reaction to go forward ([link]). In addition, an enzyme itself is unchanged by the reaction it catalyzes. Once one reaction has been catalyzed, the enzyme is able to participate in other reactions.



Enzymes lower the activation energy of the reaction but do not change the free energy of the reaction.

The chemical reactants to which an enzyme binds are called the enzyme's **substrates**. There may be one or more substrates, depending on the particular chemical reaction. In some reactions, a single reactant substrate is broken down into multiple products. In others, two substrates may come together to create one larger molecule. Two reactants might also enter a reaction and both become modified, but they leave the reaction as two products. The location within the enzyme where the substrate binds is called the enzyme's **active site**. The active site is where the "action" happens. Since enzymes are proteins, there is a unique combination of amino acid side chains within the active site. Each side chain is characterized by different properties. They can be large or small, weakly acidic or basic, hydrophilic or hydrophobic, positively or negatively charged, or neutral. The unique combination of side chains creates a very specific chemical environment within the active site. This specific environment is suited to bind to one specific chemical substrate (or substrates).

Active sites are subject to influences of the local environment. Increasing the environmental temperature generally increases reaction rates, enzyme-catalyzed or otherwise. However, temperatures outside of an optimal range reduce the rate at which an enzyme catalyzes a reaction. Hot temperatures will eventually cause enzymes to denature, an irreversible change in the three-dimensional shape and therefore the function of the enzyme. Enzymes are also suited to function best within a certain pH and salt concentration range, and, as with temperature, extreme pH, and salt concentrations can cause enzymes to denature.

For many years, scientists thought that enzyme-substrate binding took place in a simple “lock and key” fashion. This model asserted that the enzyme and substrate fit together perfectly in one instantaneous step. However, current research supports a model called induced fit ([link]). The induced-fit model expands on the lock-and-key model by describing a more dynamic binding between enzyme and substrate. As the enzyme and substrate come together, their interaction causes a mild shift in the enzyme’s structure that forms an ideal binding arrangement between enzyme and substrate.

Note:
Concept in Action

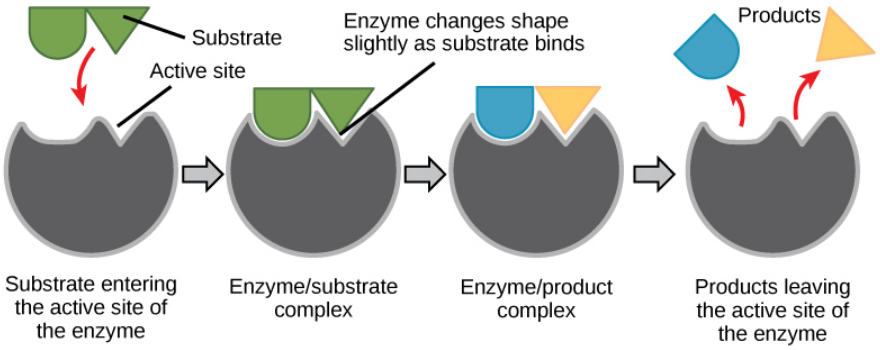


View an animation of induced fit.

When an enzyme binds its substrate, an enzyme-substrate complex is formed. This complex lowers the activation energy of the reaction and promotes its rapid progression in one of multiple possible ways. On a basic level, enzymes promote chemical reactions that involve more than one

substrate by bringing the substrates together in an optimal orientation for reaction. Another way in which enzymes promote the reaction of their substrates is by creating an optimal environment within the active site for the reaction to occur. The chemical properties that emerge from the particular arrangement of amino acid R groups within an active site create the perfect environment for an enzyme's specific substrates to react.

The enzyme-substrate complex can also lower activation energy by compromising the bond structure so that it is easier to break. Finally, enzymes can also lower activation energies by taking part in the chemical reaction itself. In these cases, it is important to remember that the enzyme will always return to its original state by the completion of the reaction. One of the hallmark properties of enzymes is that they remain ultimately unchanged by the reactions they catalyze. After an enzyme has catalyzed a reaction, it releases its product(s) and can catalyze a new reaction.



The induced-fit model is an adjustment to the lock-and-key model and explains how enzymes and substrates undergo dynamic modifications during the transition state to increase the affinity of the substrate for the active site.

It would seem ideal to have a scenario in which all of an organism's enzymes existed in abundant supply and functioned optimally under all cellular conditions, in all cells, at all times. However, a variety of

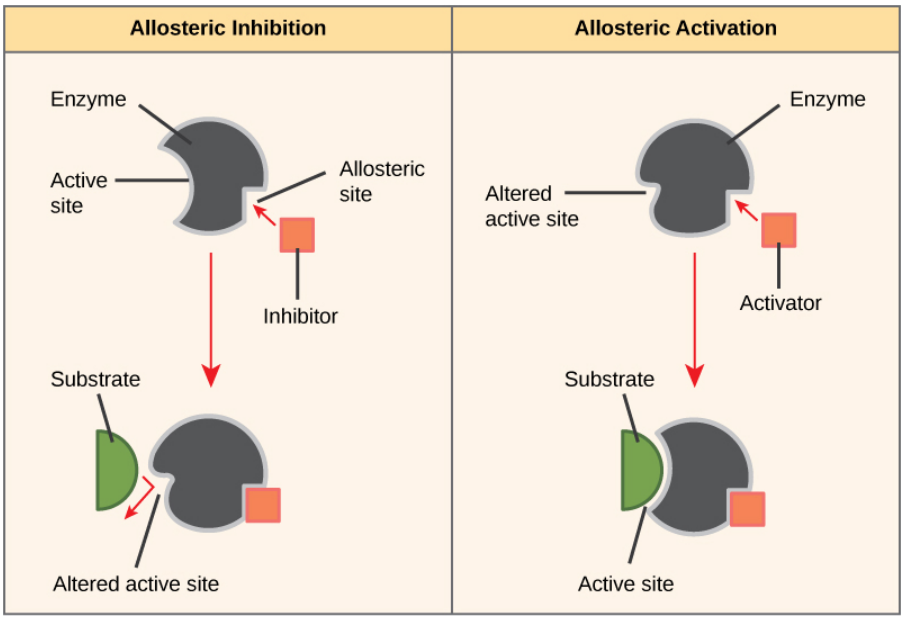
mechanisms ensures that this does not happen. Cellular needs and conditions constantly vary from cell to cell, and change within individual cells over time. The required enzymes of stomach cells differ from those of fat storage cells, skin cells, blood cells, and nerve cells. Furthermore, a digestive organ cell works much harder to process and break down nutrients during the time that closely follows a meal compared with many hours after a meal. As these cellular demands and conditions vary, so must the amounts and functionality of different enzymes.

Since the rates of biochemical reactions are controlled by activation energy, and enzymes lower and determine activation energies for chemical reactions, the relative amounts and functioning of the variety of enzymes within a cell ultimately determine which reactions will proceed and at what rates. This determination is tightly controlled in cells. In certain cellular environments, enzyme activity is partly controlled by environmental factors like pH, temperature, salt concentration, and, in some cases, cofactors or coenzymes.

Enzymes can also be regulated in ways that either promote or reduce enzyme activity. There are many kinds of molecules that inhibit or promote enzyme function, and various mechanisms by which they do so. In some cases of enzyme inhibition, an inhibitor molecule is similar enough to a substrate that it can bind to the active site and simply block the substrate from binding. When this happens, the enzyme is inhibited through **competitive inhibition**, because an inhibitor molecule competes with the substrate for binding to the active site.

On the other hand, in **noncompetitive inhibition**, an inhibitor molecule binds to the enzyme in a location other than the active site, called an allosteric site, but still manages to block substrate binding to the active site. Some inhibitor molecules bind to enzymes in a location where their binding induces a conformational change that reduces the affinity of the enzyme for its substrate. This type of inhibition is called **allosteric inhibition** ([link]). Most allosterically regulated enzymes are made up of more than one polypeptide, meaning that they have more than one protein subunit. When an allosteric inhibitor binds to a region on an enzyme, all active sites on the protein subunits are changed slightly such that they bind their substrates

with less efficiency. There are allosteric activators as well as inhibitors. Allosteric activators bind to locations on an enzyme away from the active site, inducing a conformational change that increases the affinity of the enzyme's active site(s) for its substrate(s) ([link]).



Allosteric inhibition works by indirectly inducing a conformational change to the active site such that the substrate no longer fits. In contrast, in allosteric activation, the activator molecule modifies the shape of the active site to allow a better fit of the substrate.

Note:
Careers in Action
Pharmaceutical Drug Developer



Have you ever wondered how pharmaceutical drugs are developed? (credit: Deborah Austin)

Enzymes are key components of metabolic pathways. Understanding how enzymes work and how they can be regulated are key principles behind the development of many of the pharmaceutical drugs on the market today. Biologists working in this field collaborate with other scientists to design drugs ([link]).

Consider statins for example—statins is the name given to one class of drugs that can reduce cholesterol levels. These compounds are inhibitors of the enzyme HMG-CoA reductase, which is the enzyme that synthesizes cholesterol from lipids in the body. By inhibiting this enzyme, the level of cholesterol synthesized in the body can be reduced. Similarly, acetaminophen, popularly marketed under the brand name Tylenol, is an inhibitor of the enzyme cyclooxygenase. While it is used to provide relief from fever and inflammation (pain), its mechanism of action is still not completely understood.

How are drugs discovered? One of the biggest challenges in drug discovery is identifying a drug target. A drug target is a molecule that is literally the target of the drug. In the case of statins, HMG-CoA reductase is the drug target. Drug targets are identified through painstaking research in the laboratory. Identifying the target alone is not enough; scientists also need to know how the target acts inside the cell and which reactions go awry in the case of disease. Once the target and the pathway are identified, then the actual process of drug design begins. In this stage, chemists and

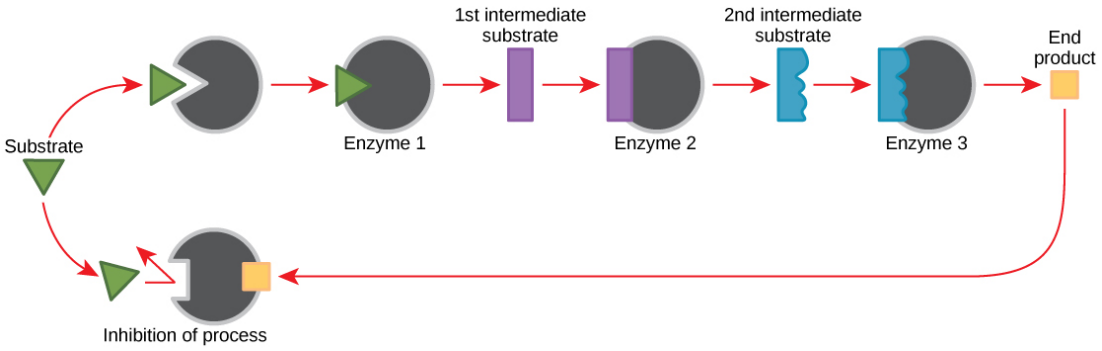
biologists work together to design and synthesize molecules that can block or activate a particular reaction. However, this is only the beginning: If and when a drug prototype is successful in performing its function, then it is subjected to many tests from in vitro experiments to clinical trials before it can get approval from the U.S. Food and Drug Administration to be on the market.

Many enzymes do not work optimally, or even at all, unless bound to other specific non-protein helper molecules. They may bond either temporarily through ionic or hydrogen bonds, or permanently through stronger covalent bonds. Binding to these molecules promotes optimal shape and function of their respective enzymes. Two examples of these types of helper molecules are cofactors and coenzymes. Cofactors are inorganic ions such as ions of iron and magnesium. Coenzymes are organic helper molecules, those with a basic atomic structure made up of carbon and hydrogen. Like enzymes, these molecules participate in reactions without being changed themselves and are ultimately recycled and reused. Vitamins are the source of coenzymes. Some vitamins are the precursors of coenzymes and others act directly as coenzymes. Vitamin C is a direct coenzyme for multiple enzymes that take part in building the important connective tissue, collagen. Therefore, enzyme function is, in part, regulated by the abundance of various cofactors and coenzymes, which may be supplied by an organism's diet or, in some cases, produced by the organism.

Feedback Inhibition in Metabolic Pathways

Molecules can regulate enzyme function in many ways. The major question remains, however: What are these molecules and where do they come from? Some are cofactors and coenzymes, as you have learned. What other molecules in the cell provide enzymatic regulation such as allosteric modulation, and competitive and non-competitive inhibition? Perhaps the most relevant sources of regulatory molecules, with respect to enzymatic cellular metabolism, are the products of the cellular metabolic reactions themselves. In a most efficient and elegant way, cells have evolved to use

the products of their own reactions for feedback inhibition of enzyme activity. **Feedback inhibition** involves the use of a reaction product to regulate its own further production ([link]). The cell responds to an abundance of the products by slowing down production during anabolic or catabolic reactions. Such reaction products may inhibit the enzymes that catalyzed their production through the mechanisms described above.



Metabolic pathways are a series of reactions catalyzed by multiple enzymes. Feedback inhibition, where the end product of the pathway inhibits an upstream process, is an important regulatory mechanism in cells.

The production of both amino acids and nucleotides is controlled through feedback inhibition. Additionally, ATP is an allosteric regulator of some of the enzymes involved in the catabolic breakdown of sugar, the process that creates ATP. In this way, when ATP is in abundant supply, the cell can prevent the production of ATP. On the other hand, ADP serves as a positive allosteric regulator (an allosteric activator) for some of the same enzymes that are inhibited by ATP. Thus, when relative levels of ADP are high compared to ATP, the cell is triggered to produce more ATP through sugar catabolism.

Section Summary

Cells perform the functions of life through various chemical reactions. A cell’s metabolism refers to the combination of chemical reactions that take place within it. Catabolic reactions break down complex chemicals into simpler ones and are associated with energy release. Anabolic processes build complex molecules out of simpler ones and require energy.

In studying energy, the term system refers to the matter and environment involved in energy transfers. Entropy is a measure of the disorder of a system. The physical laws that describe the transfer of energy are the laws of thermodynamics. The first law states that the total amount of energy in the universe is constant. The second law of thermodynamics states that every energy transfer involves some loss of energy in an unusable form, such as heat energy. Energy comes in different forms: kinetic, potential, and free. The change in free energy of a reaction can be negative (releases energy, exergonic) or positive (consumes energy, endergonic). All reactions require an initial input of energy to proceed, called the activation energy.

Enzymes are chemical catalysts that speed up chemical reactions by lowering their activation energy. Enzymes have an active site with a unique chemical environment that fits particular chemical reactants for that enzyme, called substrates. Enzymes and substrates are thought to bind according to an induced-fit model. Enzyme action is regulated to conserve resources and respond optimally to the environment.

Art Connections

Exercise:

Problem:

[link] Look at each of the processes shown and decide if it is endergonic or exergonic.

Solution:

[link] A compost pile decomposing is an exergonic process. A baby developing from a fertilized egg is an endergonic process. Tea

dissolving into water is an exergonic process. A ball rolling downhill is an exergonic process.

Review Questions

Exercise:

Problem:

Which of the following is not an example of an energy transformation?

a. Heating up dinner in a microwave
b. Solar panels at work
c. Formation of static electricity
d. None of the above

Solution:

D

Exercise:

Problem: Which of the following is not true about enzymes?

a. They are consumed by the reactions they catalyze.
b. They are usually made of amino acids.
c. They lower the activation energy of chemical reactions.
d. Each one is specific to the particular substrate(s) to which it binds.

Solution:

A

Free Response

Exercise:

Problem:

Does physical exercise to increase muscle mass involve anabolic and/or catabolic processes? Give evidence for your answer.

Solution:

Physical exercise involves both anabolic and catabolic processes. Body cells break down sugars to provide ATP to do the work necessary for exercise, such as muscle contractions. This is catabolism. Muscle cells also must repair muscle tissue damaged by exercise by building new muscle. This is anabolism.

Exercise:

Problem:

Explain in your own terms the difference between a spontaneous reaction and one that occurs instantaneously, and what causes this difference.

Solution:

A spontaneous reaction is one that has a negative ΔG and thus releases energy. However, a spontaneous reaction need not occur quickly or suddenly like an instantaneous reaction. It may occur over long periods of time due to a large energy of activation, which prevents the reaction from occurring quickly.

Exercise:

Problem:

With regard to enzymes, why are vitamins and minerals necessary for good health? Give examples.

Solution:

Most vitamins and minerals act as cofactors and coenzymes for enzyme action. Many enzymes require the binding of certain cofactors or coenzymes to be able to catalyze their reactions. Since enzymes catalyze many important reactions, it is critical to obtain sufficient vitamins and minerals from diet and supplements. Vitamin C (ascorbic acid) is a coenzyme necessary for the action of enzymes that build collagen.

Glossary

activation energy
: the amount of initial energy necessary for reactions to occur

active site
: a specific region on the enzyme where the substrate binds

allosteric inhibition
: the mechanism for inhibiting enzyme action in which a regulatory molecule binds to a second site (not the active site) and initiates a conformation change in the active site, preventing binding with the substrate

anabolic
: describes the pathway that requires a net energy input to synthesize complex molecules from simpler ones

bioenergetics
: the concept of energy flow through living systems

catabolic
: describes the pathway in which complex molecules are broken down into simpler ones, yielding energy as an additional product of the reaction

competitive inhibition
: a general mechanism of enzyme activity regulation in which a molecule other than the enzyme’s substrate is able to bind the active

site and prevent the substrate itself from binding, thus inhibiting the overall rate of reaction for the enzyme

endergonic
: describes a chemical reaction that results in products that store more chemical potential energy than the reactants

enzyme
: a molecule that catalyzes a biochemical reaction

exergonic
: describes a chemical reaction that results in products with less chemical potential energy than the reactants, plus the release of free energy

feedback inhibition
: a mechanism of enzyme activity regulation in which the product of a reaction or the final product of a series of sequential reactions inhibits an enzyme for an earlier step in the reaction series

heat energy
: the energy transferred from one system to another that is not work

kinetic energy
: the type of energy associated with objects in motion

metabolism
: all the chemical reactions that take place inside cells, including those that use energy and those that release energy

noncompetitive inhibition
: a general mechanism of enzyme activity regulation in which a regulatory molecule binds to a site other than the active site and prevents the active site from binding the substrate; thus, the inhibitor molecule does not compete with the substrate for the active site; allosteric inhibition is a form of noncompetitive inhibition

potential energy

the type of energy that refers to the potential to do work

substrate
: a molecule on which the enzyme acts

thermodynamics
: the science of the relationships between heat, energy, and work

Energy in Living Systems
By the end of this section, you will be able to:

- Discuss the importance of electrons in the transfer of energy in living systems
- Explain how ATP is used by the cell as an energy source

Energy production within a cell involves many coordinated chemical pathways. Most of these pathways are combinations of oxidation and reduction reactions. Oxidation and reduction occur in tandem. An oxidation reaction strips an electron from an atom in a compound, and the addition of this electron to another compound is a reduction reaction. Because oxidation and reduction usually occur together, these pairs of reactions are called oxidation reduction reactions, or **redox reactions**.

Electrons and Energy

The removal of an electron from a molecule, oxidizing it, results in a decrease in potential energy in the oxidized compound. The electron (sometimes as part of a hydrogen atom), does not remain unbonded, however, in the cytoplasm of a cell. Rather, the electron is shifted to a second compound, reducing the second compound. The shift of an electron from one compound to another removes some potential energy from the first compound (the oxidized compound) and increases the potential energy of the second compound (the reduced compound). The transfer of electrons between molecules is important because most of the energy stored in atoms and used to fuel cell functions is in the form of high-energy electrons. The transfer of energy in the form of electrons allows the cell to transfer and use energy in an incremental fashion—in small packages rather than in a single, destructive burst. This chapter focuses on the extraction of energy from food; you will see that as you track the path of the transfers, you are tracking the path of electrons moving through metabolic pathways.

Electron Carriers

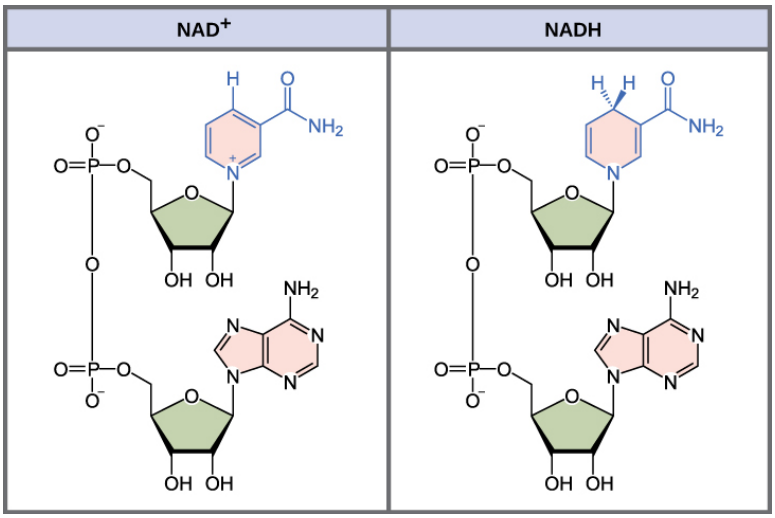
In living systems, a small class of compounds functions as electron shuttles: They bind and carry high-energy electrons between compounds in pathways. The principal electron carriers we will consider are derived from the B vitamin group and are derivatives of nucleotides. These compounds can be easily reduced (that is, they accept electrons) or oxidized (they lose electrons). Nicotinamide adenine dinucleotide (NAD) ([link]) is derived from vitamin B3, niacin. NAD^+ is the oxidized form of the molecule; NADH is the reduced form of the molecule after it has accepted two electrons and a proton (which together are the equivalent of a hydrogen atom with an extra electron).

NAD^+ can accept electrons from an organic molecule according to the general equation:
Equation:

$$\underset{\text{Reducing agent}}{RH} + \underset{\text{Oxidizing agent}}{NAD^+} \rightarrow \underset{\text{Reduced}}{NADH} + \underset{\text{Oxidized}}{R}$$

When electrons are added to a compound, they are reduced. A compound that reduces another is called a reducing agent. In the above equation, RH is a reducing agent, and NAD^+ is reduced to NADH. When electrons are removed from compound, it oxidized. A compound that oxidizes another is called an oxidizing agent. In the above equation, NAD^+ is an oxidizing agent, and RH is oxidized to R.

Similarly, flavin adenine dinucleotide (FAD^+) is derived from vitamin B_2, also called riboflavin. Its reduced form is $FADH_2$. A second variation of NAD, NADP, contains an extra phosphate group. Both NAD^+ and FAD^+ are extensively used in energy extraction from sugars, and NADP plays an important role in anabolic reactions and photosynthesis.



The oxidized form of the electron carrier (NAD^+) is shown on the left and the reduced form (NADH) is shown on the right. The nitrogenous base in NADH has one more hydrogen ion and two more electrons than in NAD^+.

ATP in Living Systems

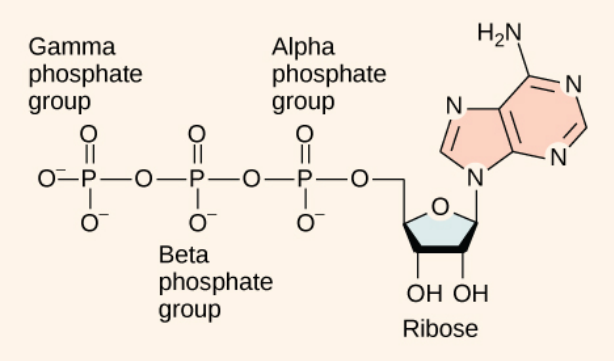
A living cell cannot store significant amounts of free energy. Excess free energy would result in an increase of heat in the cell, which would result in excessive thermal motion that could damage and then destroy the cell. Rather, a cell must be able to handle that energy in a way that enables the cell to store energy safely and release it for use only as needed. Living cells accomplish this by using the compound adenosine triphosphate (ATP). ATP is often called the “energy currency” of the cell, and, like currency, this versatile compound can be used to fill any energy need of the cell. How? It functions similarly to a rechargeable battery.

When ATP is broken down, usually by the removal of its terminal phosphate group, energy is released. The energy is used to do work by the cell, usually by the released phosphate binding to

another molecule, activating it. For example, in the mechanical work of muscle contraction, ATP supplies the energy to move the contractile muscle proteins. Recall the active transport work of the sodium-potassium pump in cell membranes. ATP alters the structure of the integral protein that functions as the pump, changing its affinity for sodium and potassium. In this way, the cell performs work, pumping ions against their electrochemical gradients.

ATP Structure and Function

At the heart of ATP is a molecule of adenosine monophosphate (AMP), which is composed of an adenine molecule bonded to a ribose molecule and to a single phosphate group ([link]). Ribose is a five-carbon sugar found in RNA, and AMP is one of the nucleotides in RNA. The addition of a second phosphate group to this core molecule results in the formation of adenosine diphosphate (ADP); the addition of a third phosphate group forms adenosine triphosphate (ATP).



ATP (adenosine triphosphate) has three phosphate groups that can be removed by hydrolysis to form ADP (adenosine diphosphate) or AMP (adenosine monophosphate).The negative charges on the phosphate group naturally repel each other, requiring energy to bond them together and releasing energy when these bonds are broken.

The addition of a phosphate group to a molecule requires energy. Phosphate groups are negatively charged and thus repel one another when they are arranged in series, as they are in ADP and ATP. This repulsion makes the ADP and ATP molecules inherently unstable. The release of one or two phosphate groups from ATP, a process called **dephosphorylation**, releases energy.

Energy from ATP

Hydrolysis is the process of breaking complex macromolecules apart. During hydrolysis, water is split, or lysed, and the resulting hydrogen atom (H^+) and a hydroxyl group (OH^-) are added to the larger molecule. The hydrolysis of ATP produces ADP, together with an inorganic phosphate ion (P_i), and the release of free energy. To carry out life processes, ATP is continuously broken down into ADP, and like a rechargeable battery, ADP is continuously regenerated into ATP by the reattachment of a third phosphate group. Water, which was broken down into its hydrogen atom and hydroxyl group during ATP hydrolysis, is regenerated when a third phosphate is added to the ADP molecule, reforming ATP.

Obviously, energy must be infused into the system to regenerate ATP. Where does this energy come from? In nearly every living thing on earth, the energy comes from the metabolism of glucose. In this way, ATP is a direct link between the limited set of exergonic pathways of glucose catabolism and the multitude of endergonic pathways that power living cells.

Phosphorylation

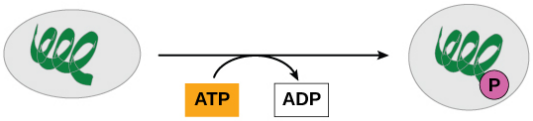
Recall that, in some chemical reactions, enzymes may bind to several substrates that react with each other on the enzyme, forming an intermediate complex. An intermediate complex is a temporary structure, and it allows one of the substrates (such as ATP) and reactants to more readily react with each other; in reactions involving ATP, ATP is one of the substrates and ADP is a product. During an endergonic chemical reaction, ATP forms an intermediate complex with the substrate and enzyme in the reaction. This intermediate complex allows the ATP to transfer its third phosphate group, with its energy, to the substrate, a process called phosphorylation.
Phosphorylation refers to the addition of the phosphate (~P). This is illustrated by the following generic reaction:
Equation:

$$\mathrm{A + enzyme + ATP \rightarrow\ [A\ -\ enzyme\ -\ \sim P]\ \rightarrow\ B + enzyme + ADP + phosphate\ ion}$$

When the intermediate complex breaks apart, the energy is used to modify the substrate and convert it into a product of the reaction. The ADP molecule and a free phosphate ion are released into the medium and are available for recycling through cell metabolism.

Substrate Phosphorylation

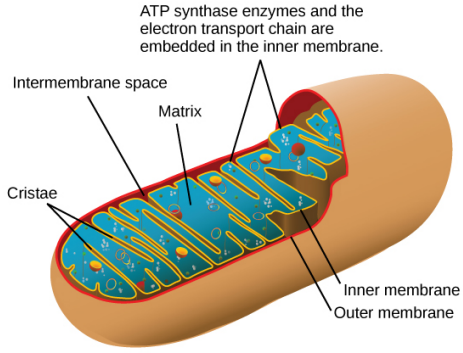
ATP is generated through two mechanisms during the breakdown of glucose. A few ATP molecules are generated (that is, regenerated from ADP) as a direct result of the chemical reactions that occur in the catabolic pathways. A phosphate group is removed from an intermediate reactant in the pathway, and the free energy of the reaction is used to add the third phosphate to an available ADP molecule, producing ATP ([link]). This very direct method of phosphorylation is called **substrate-level phosphorylation**.



In phosphorylation reactions, the gamma phosphate of ATP is attached to a protein.

Oxidative Phosphorylation

Most of the ATP generated during glucose catabolism, however, is derived from a much more complex process, chemiosmosis, which takes place in mitochondria ([link]) within a eukaryotic cell or the plasma membrane of a prokaryotic cell. **Chemiosmosis**, a process of ATP production in cellular metabolism, is used to generate 90 percent of the ATP made during glucose catabolism and is also the method used in the light reactions of photosynthesis to harness the energy of sunlight. The production of ATP using the process of chemiosmosis is called **oxidative phosphorylation** because of the involvement of oxygen in the process.



In eukaryotes, oxidative phosphorylation takes place in mitochondria. In prokaryotes, this process takes place in the plasma membrane. (Credit: modification of work by Mariana Ruiz Villareal)

Note:

Career Connections

Mitochondrial Disease Physician

What happens when the critical reactions of cellular respiration do not proceed correctly? Mitochondrial diseases are genetic disorders of metabolism. Mitochondrial disorders can arise from mutations in nuclear or mitochondrial DNA, and they result in the production of less energy than is normal in body cells. In type 2 diabetes, for instance, the oxidation efficiency of NADH is reduced, impacting oxidative phosphorylation but not the other steps of respiration. Symptoms of mitochondrial diseases can include muscle weakness, lack of coordination, stroke-like episodes, and loss of vision and hearing. Most affected people are diagnosed in childhood, although there are some adult-onset diseases. Identifying and treating mitochondrial disorders is a specialized medical field. The educational preparation for this profession requires a college education, followed by medical school with a specialization in medical genetics. Medical geneticists can be board certified by the American Board of Medical Genetics and go on to become associated with professional organizations devoted to the study of mitochondrial diseases, such as the Mitochondrial Medicine Society and the Society for Inherited Metabolic Disease.

Section Summary

ATP functions as the energy currency for cells. It allows the cell to store energy briefly and transport it within the cell to support endergonic chemical reactions. The structure of ATP is that of an RNA nucleotide with three phosphates attached. As ATP is used for energy, a phosphate group or two are detached, and either ADP or AMP is produced. Energy derived from glucose catabolism is used to convert ADP into ATP. When ATP is used in a reaction, the third phosphate is temporarily attached to a substrate in a process called phosphorylation. The two processes of ATP regeneration that are used in conjunction with glucose catabolism are substrate-level phosphorylation and oxidative phosphorylation through the process of chemiosmosis.

Review Questions

Exercise:

Problem: The energy currency used by cells is _____.

a. ATP
b. ADP
c. AMP
d. adenosine

Solution:

A

Exercise:

Problem: A reducing chemical reaction _____.

a. reduces the compound to a simpler form
b. adds an electron to the substrate
c. removes a hydrogen atom from the substrate
d. is a catabolic reaction

Solution:

B

Free Response

Exercise:

Problem:

Why is it beneficial for cells to use ATP rather than energy directly from the bonds of carbohydrates? What are the greatest drawbacks to harnessing energy directly from the bonds of several different compounds?

Solution:

ATP provides the cell with a way to handle energy in an efficient manner. The molecule can be charged, stored, and used as needed. Moreover, the energy from hydrolyzing ATP is delivered as a consistent amount. Harvesting energy from the bonds of several different compounds would result in energy deliveries of different quantities.

Glossary

chemiosmosis
: process in which there is a production of adenosine triphosphate (ATP) in cellular metabolism by the involvement of a proton gradient across a membrane

dephosphorylation
: removal of a phosphate group from a molecule

oxidative phosphorylation
: production of ATP using the process of chemiosmosis and oxygen

phosphorylation
: addition of a high-energy phosphate to a compound, usually a metabolic intermediate, a protein, or ADP

redox reaction
: chemical reaction that consists of the coupling of an oxidation reaction and a reduction reaction

substrate-level phosphorylation

production of ATP from ADP using the excess energy from a chemical reaction and a phosphate group from a reactant

Bis2A 06.0 Energy in Living Systems v1.2
By the end of this section, you will be able to:

- Discuss the importance of electrons in the transfer of energy in living systems
- Explain how ATP is used by the cell as an energy source

Energy production within a cell involves many coordinated chemical pathways. Most of these pathways are combinations of oxidation and reduction reactions. Oxidation and reduction occur in tandem. An oxidation reaction strips an electron from an atom in a compound, and the addition of this electron to another compound is a reduction reaction. Because oxidation and reduction usually occur together, these pairs of reactions are called oxidation reduction reactions, or **redox reactions**.

Oxidation-Reduction Reactions

The chemical reactions underlying metabolism involve the transfer of electrons from one compound to another by processes catalyzed by enzymes. The electrons in these reactions commonly come from hydrogen atoms, which consist of an electron and a proton. A molecule gives up a hydrogen atom, in the form of a hydrogen ion (H^+) and an electron, breaking the molecule into smaller parts. The loss of an electron, or **oxidation**, releases a small amount of energy; both the electron and the energy are then passed to another molecule in the process of **reduction**, or the gaining of an electron. These two reactions always happen together in an **oxidation-reduction reaction** (also called a redox reaction)—when an electron is passed between molecules, the donor is oxidized and the recipient is reduced. Oxidation-reduction reactions often happen in a series, so that a molecule that is reduced is subsequently oxidized, passing on not only the electron it just received but also the energy it received. As the series of reactions progresses, energy accumulates that is used to combine P_i and ADP to form ATP, the high-energy molecule that the body uses for fuel.

Oxidation-reduction reactions are catalyzed by enzymes that trigger the removal of electrons (either one or two) from a substrate (sometimes the removal coincides with the removal of a proton) and transfers then to a second substrate, usually a coenzyme which can temporarily maintain the electrons (and sometimes protons) before transferring then to a second compound. There are two broad classes of coenzymes that work in redox reactions in the cell. The first are those coenzymesm that can carry both electrons and protons, the two most common are **nicotinamide adenine dinucleotide (NAD^+)** and **flavin adenine dinucleotide (FAD)**. Their respective reduced coenzymes are **NADH** and **$FADH_2$**. A third coenzyme, **nicotinamide adenine dinucleotide phosphate ($NADP^+$)** is the primary reductant for **anabolic** reactions and has similar (yet distinct) properties to its unphosphorylated counterpart **NAD^+**. The second general group of enzymes (that contain co-factors such as heme) only carry electrons, and include **cytochromes**, and **Iron-Sulfur (Fe-S) proteins**.
Exercise:
Review Question

Problem:

Consider a red/ox reaction that requires NAD as a co-enzyme. As the reaction proceeds NAD+ is reduced to NADH. What would happen to the reaction rate and the substrate concentration if the NAD+ pool is fixed (a finite number of molecules in the cell)?

Solution:

As the reaction proceeds, NAD+ is reduced to NADH, if the NAD+ concentration is fixed, and NADH is not recycled the reaction will stop (or sit at some equilibrium)with no net increase in NADH because the NAD+ pool becomes so low.

Electrons and Energy

The removal of an electron from a molecule, oxidizing it, results in a decrease in potential energy in the oxidized compound. The electron (sometimes as part of a hydrogen atom), does not remain unbonded, however, in the cytoplasm of a cell. Rather, the electron is shifted to a second compound, reducing the second compound. The shift of an electron from one compound to another removes some potential energy from the first compound (the oxidized compound) and increases the potential energy of the second compound (the reduced compound). The transfer of electrons between molecules is important because most of the energy stored in atoms and used to fuel cell functions is in the form of high-energy electrons. The transfer of energy in the form of electrons allows the cell to transfer and use energy in an incremental fashion—in small packages rather than in a single, destructive burst. This chapter focuses on the extraction of energy from food; you will see that as you track the path of the transfers, you are tracking the path of electrons moving through metabolic pathways.

Electron Carriers

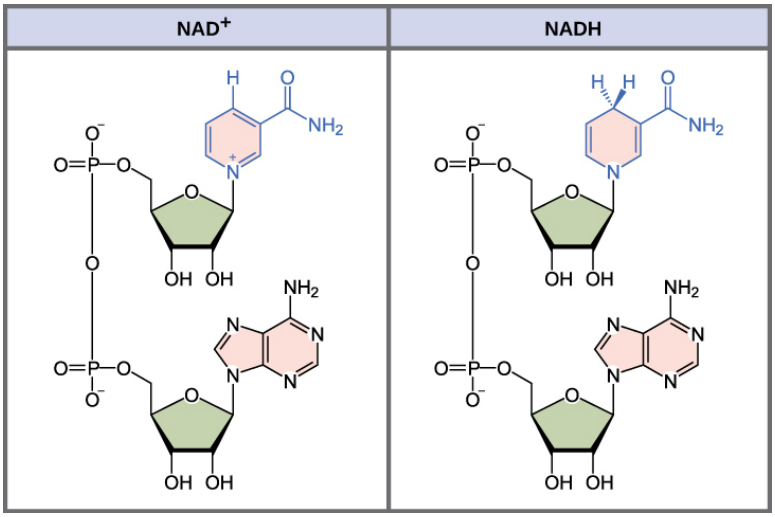
In living systems, a small class of compounds functions as electron shuttles: They bind and carry high-energy electrons between compounds in pathways. The principal electron carriers we will consider are derived from the B vitamin group and are derivatives of nucleotides. These compounds can be easily reduced (that is, they accept electrons) or oxidized (they lose electrons). Nicotinamide adenine dinucleotide (NAD+) ([link]) is derived from vitamin B3, niacin. NAD^+ is the oxidized form of the molecule; NADH is the reduced form of the molecule after it has accepted two electrons and a proton (which together are the equivalent of a hydrogen atom with an extra electron).

NAD^+ can accept electrons from an organic molecule according to the general equation:
Equation:

$$\underset{\text{Reducing agent}}{\text{RH}} + \underset{\text{Oxidizing agent}}{\text{NAD}^+} \rightarrow \underset{\text{Reduced}}{\text{NADH}} + \underset{\text{Oxidized}}{\text{R}}$$

When electrons are added to a compound, they are reduced. A compound that reduces another is called a reducing agent. In the above equation, RH is a reducing agent, and NAD^+ is reduced to NADH. When electrons are removed from compound, it oxidized. A compound that oxidizes another is called an oxidizing agent. In the above equation, NAD^+ is an oxidizing agent, and RH is oxidized to R.

Similarly, flavin adenine dinucleotide (FAD^+) is derived from vitamin B_2, also called riboflavin. Its reduced form is $FADH_2$. A second variation of NAD, NADP, contains an extra phosphate group. Both NAD^+ and FAD^+ are extensively used in energy extraction from sugars, and NADP plays an important role in anabolic reactions and photosynthesis.



The oxidized form of the electron carrier (NAD^+) is shown on the left and the reduced form (NADH) is shown on the right. The nitrogenous base in NADH has one more hydrogen ion and two more electrons than in NAD^+.

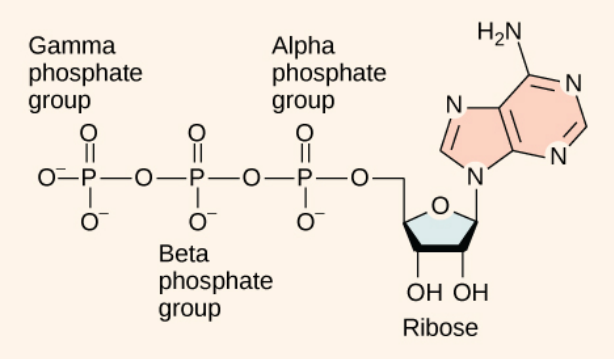
ATP in Living Systems

A living cell cannot store significant amounts of free energy. Excess free energy would result in an increase of heat in the cell, which would result in excessive thermal motion that could damage and then destroy the cell. Rather, a cell must be able to handle that energy in a way that enables the cell to store energy safely and release it for use only as needed. Living cells accomplish this by using the compound adenosine triphosphate (ATP). ATP is often called the “energy currency” of the cell, and, like currency, this versatile compound can be used to fill any energy need of the cell. How? It functions similarly to a rechargeable battery.

When ATP is broken down, usually by the removal of its terminal phosphate group, energy is released. The energy is used to do work by the cell, usually by the released phosphate binding to another molecule, activating it. For example, in the mechanical work of muscle contraction, ATP supplies the energy to move the contractile muscle proteins. Recall the active transport work of the sodium-potassium pump in cell membranes. ATP alters the structure of the integral protein that functions as the pump, changing its affinity for sodium and potassium. In this way, the cell performs work, pumping ions against their electrochemical gradients.

ATP Structure and Function

At the heart of ATP is a molecule of adenosine monophosphate (AMP), which is composed of an adenine molecule bonded to a ribose molecule and to a single phosphate group ([link]). Ribose is a five-carbon sugar found in RNA, and AMP is one of the nucleotides in RNA. The addition of a second phosphate group to this core molecule results in the formation of adenosine diphosphate (ADP); the addition of a third phosphate group forms adenosine triphosphate (ATP).



ATP (adenosine triphosphate) has three phosphate groups that can be removed by hydrolysis to form ADP (adenosine diphosphate) or AMP (adenosine monophosphate).The negative charges on the phosphate group naturally repel each other, requiring energy to bond them together and releasing energy when these bonds are broken.

The addition of a phosphate group to a molecule requires energy. Phosphate groups are negatively charged and thus repel one another when they are arranged in series, as they are in ADP and ATP. This repulsion makes the ADP and ATP molecules inherently unstable. The release of one or two phosphate groups from ATP, a process called **dephosphorylation**, releases energy.

Video on electron and electron/proton carriers
For a 7 minute YouTube video on the role of carriers in respiration clicke here.

Energy from ATP

Hydrolysis is the process of breaking complex macromolecules apart. During hydrolysis, water is split, or lysed, and the resulting hydrogen atom (H^+) and a hydroxyl group (OH^-) are added to the larger molecule. The hydrolysis of ATP produces ADP, together with an inorganic phosphate ion (P_i), and the release of free energy. To carry out life processes, ATP is continuously broken down

into ADP, and like a rechargeable battery, ADP is continuously regenerated into ATP by the reattachment of a third phosphate group. Water, which was broken down into its hydrogen atom and hydroxyl group during ATP hydrolysis, is regenerated when a third phosphate is added to the ADP molecule, reforming ATP.

Obviously, energy must be infused into the system to regenerate ATP. Where does this energy come from? In nearly every living thing on earth, the energy comes from the metabolism of glucose. In this way, ATP is a direct link between the limited set of exergonic pathways of glucose catabolism and the multitude of endergonic pathways that power living cells.

Video Link
For a more detailed explanation of ATP and how this small molecule can store so much energy, take a look at this video (10 minutes) by clicking here.

Phosphorylation

Recall that, in some chemical reactions, enzymes may bind to several substrates that react with each other on the enzyme, forming an intermediate complex. An intermediate complex is a temporary structure, and it allows one of the substrates (such as ATP) and reactants to more readily react with each other; in reactions involving ATP, ATP is one of the substrates and ADP is a product. During an endergonic chemical reaction, ATP forms an intermediate complex with the substrate and enzyme in the reaction. This intermediate complex allows the ATP to transfer its third phosphate group, with its energy, to the substrate, a process called phosphorylation.
Phosphorylation refers to the addition of the phosphate (~P). This is illustrated by the following generic reaction:
Equation:

$$\mathrm{A} + \mathrm{enzyme} + \mathrm{ATP} \rightarrow \ [\mathrm{A} \ - \ \mathrm{enzyme} \ - \ \sim \mathrm{P}] \ \rightarrow \ \mathrm{B} + \mathrm{enzyme} + \mathrm{ADP} + \mathrm{phosphate\ ion}$$

When the intermediate complex breaks apart, the energy is used to modify the substrate and convert it into a product of the reaction. The ADP molecule and a free phosphate ion are released into the medium and are available for recycling through cell metabolism.

How do cells generate ATP

All cells require energy in the form of mobile packets. As we saw earlier in the class, these packets can be used to couple thermodynamically unfavorable reactions to drive the formation of specific products. Remember it costs energy to build, and the primary role of a cell is to make two cells. For what ever reason, almost 3.25 billion years of evolution has favored **ATP** as that mobile form of energy. It is the energy held within the **phosphoanhydride** bonds (the terminal or gamma and beta Phosphates) that store the energy. As shown in Figure 3, many common cellular compounds have such bonds and can also be used as an internal energy source with specific enzymes. Such compounds include all of the NTPs as well as common intermediates such as Phosphoenol pyruvate (PEP), which we will discuss later in glycolysis. Remember Not all carbon-Phosphate bonds are high energy, such as glucose-6-phosphate (figure 3).

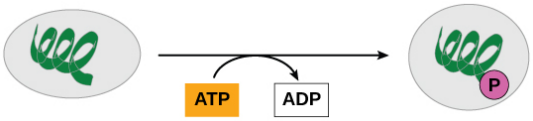
Compound	$\Delta G^{0'}$ kJ/mol
> $\Delta G^{0'}$ kJ/mol	
Phosphoenol pyruvate	-51.6
1,3-bis-Phosphoglycerate	-52.0
Acetylphosphate	-44.8
ATP	-31.8
ADP	-31.8
Acetyl~CoA	-31
< $\Delta G^{0'}$ kJ/mol	
AMP	-14.2
Glucose-6-P	-13.8

Table of common cellular phosphorylated molecules and there respective free energies

Because ATP is the primary source of mobile energy in the cell, a variety of mechanisms have ermerged over the 3.25 billion years of evolution to form ATP. The majority of these mechanism are modifications on two themes: direct synthesis of ATP or indirect synthesis of ATP. However all of the known reactions fall into two basicmechanisms: **Substrate Level Phosphorylation (SLP)** and **oxidative phosphorylation** and will be discussed in detail below and in the next few modules. Suffice it to say both mechanisms rely on the transfer of potential energy from one high energy compound (often called the energy source) to ADP, to synthesize ATP. This is either done directly, as in the case of Substrate level phosphorylation, or indirectly, as is the case for oxidative phosphorylation.

Substrate Level Phosphorylation

The simplest rout to synthesize ATP is substrate level phosphorylation. ATP molecules are generated (that is, regenerated from ADP) as a direct result of a chemical reaction that occurs in catabolic pathways. A phosphate group is removed from an intermediate reactant in the pathway, and the free energy of the reaction is used to add the third phosphate to an available ADP molecule, producing ATP ([link]). This very direct method of phosphorylation is called **substrate-level phosphorylation**. It can be found in a variety of catabolic reactions, most notably in two specific reactions in glycolysis (which we will discuss specifically later). Suffice it to say what is required is a high energy intermediate whose oxidation is sufficient to drive the synthesis of ATP.

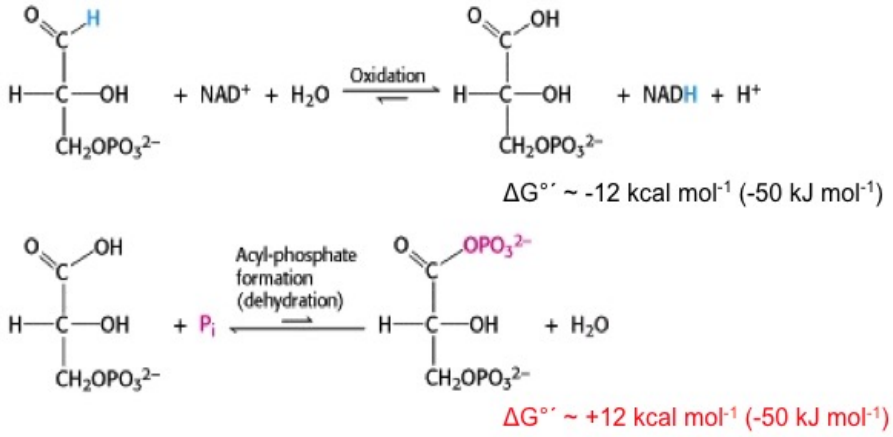


In phosphorylation reactions, the gamma phosphate of ATP is attached to a protein.

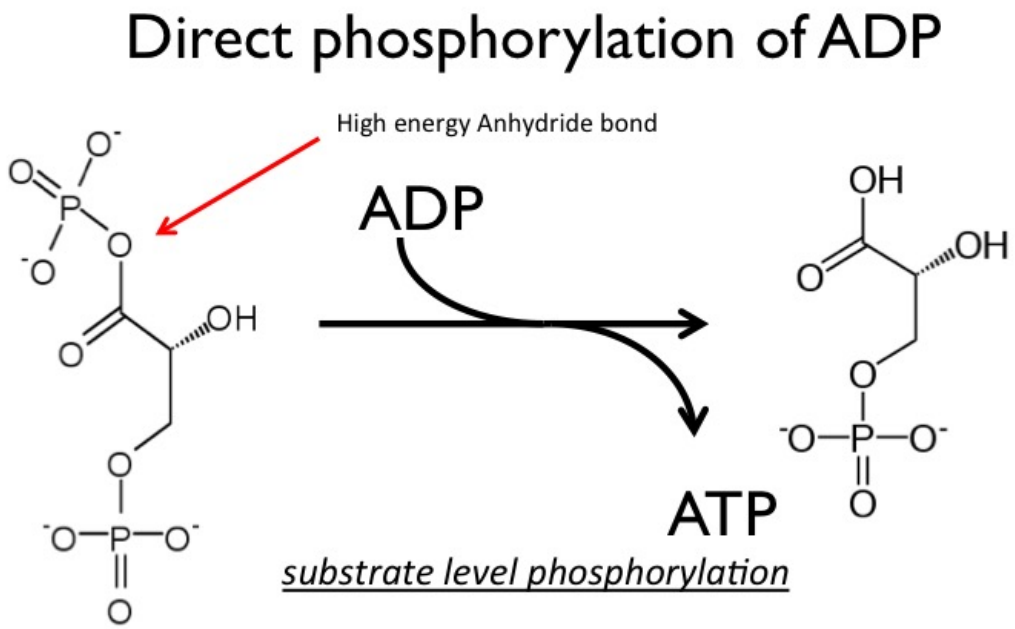
From an evolutionary perspective consider remember that the primordial ooze was hypothesized to be enriched with highly reduced small organic compounds. These reduced high energy compounds could have made an excellent initial energy source for early life. The simplest method of energy extraction being the oxidation of these compounds coupled to ATP synthesis; i.e. Substrate level phosphorylation. This simple method could theoretically have yielded large quantities of ATP for the cell along with smaller more oxidized organic compounds. Remember, we are currently only discussing ATP synthesis. If you consider this reaction, the coupling of the oxidation of reduced orgainic molecules to ATP synthesis, we are missing half the reaction, the compound to be reduced; ATP is not reduced during this reaction, it is simply forming a high energy bond. The electrons need to go somewhere, and the question where do they go?

The short answer is NAD+, those co-enzymes involved in red/ox reactions. Thus the first step in the synthesis of ATP is drive a Phosphate group onto the high energy compound by a red/ox reaction. The subsequent reduction of the energy source with the simultaneous phosphorylation produces two products, NADH (from the reduction of NAD+)and the newly formed phosphorylated compound containing a high energy phosphate that can now be used to synthesize ATP in a second reaction. The best example of this in current metabolism is the reduction of glyceraldehyde-3-phosphate to 1,3-bisposphoglycerate by NAD+. The formation of 1,3-bisphosphoglycerate and a high energy phosphoanhydride can then be used to phosphorylate ADP to ATP and production of 3-phosphoglycerate. This is diagrammed in figure 5 below.

The oxidation of simultaneous phosphorylation of glyceraldehyde-3-phosphate:
This can be thought of as 2 reactions: Red/Ox and phosphorylation



The subsequent phosphate transfer reaction to ADP to form ATP in the reaction shown in figure 6 comes directly from the formation of 1,3-bisphosphoglycerate (described above). The arrow points out the high energy phosphoanhydride that is formed in the red/ox reaction. These two enzymatic reactions demonstrate how cells can transform one form of energy into a second, the potential energy from the substrates to potential energy in ATP. And of course the energy in ATP can then be used to help drive thermodynamically unfavorable reactions in the cell.



There is one last consideration you need to think about in our discussion of SLP. Consider the question posed earlier in exercise 1. Remember NAD+ is the original oxidizing agent used to form 1,3-bisphosphoglycerate from glyceraldehyde-3-phosphate. NAD+ is a co-enzyme and its levels are finite in the cell. So, as the reactions in figures 5 and 6 occur in the cell the levels of NAD+ fall and the levels of NADH rise. If left unchecked the cell has entered a death spiral. The answer is relatively simple, regenerate NAD+ by oxidizing NADH. This process is called **fermentation** and its role in metabolism is to reoxidize NADH to NAD+ forming a cycle between the two forms of the same molecule; as long as there is plenty of substrate to reduce NAD+ and enough substrate to oxidize NADH back to NAD+. We will discus fermentation reactions in detail in a different module. But for this discussion, simply keep in mind that the cell must reoxidize NADH back to NAD+ in order to survive. Also keep in mind, that as with all aspects of energy metabolism, the basic process is very simple, NAD+ to NADH to NAD+ to....., yet a variety of rich complex reactions have evolved over time. This good news for us, because the end products of fermentation reactions include many foods and beverages we use every day. Think how boring life would be without bread, tea, beer and chocolate.

Oxidative Phosphorylation

In addition to substrate level phosphorylation, cells can generate ATP via an **electron transport chain** coupled to the proton dependent F0F1 ATPase. Such coupled systems require a **membrane**to both separate charge (electrical potential) and protons (chemical potential). This process is refereed, called the **Chemiosmotic hypothesis** was proposed by Peter Mitchell in 1961.

While we will discuss this in detail later, in essence oxidative phosphorylation works by taking a reduced high energy compound, such as NADH+H+ or FADH2, and transferring electrons in a

series of redox reactions to other compounds. The energy that is release is then "captured" by the translocation of protons (H+) across the membrane. The net result is an "energize membrand" that separates charge, negative on the inside and positive on the outside (the electrical potential) and the separation of protons (chemical potential), protons concentration greater on the outside then inside of the membrane. The overall energized state of the membrane is then similar to a charged capacitor ready to do work. ATP generation occurs by the F0F1 membrane bound ATPase that translocates protons across the membrane discharging the chemical and electrical potentials with the simultaneous synthesis of ATP. For every three protons (H+) translocated, the F0F1ATPase can synthesize 1 ATP.

Finally, oxidative phosphorylation does NOT require oxygen to function. The word oxidative in this case is in reference to redox reactions driving the process. We are very familiar with aerobic respiration which uses molecular oxygen as the terminal electron acceptor in some electron transport chains but this is a subset of reactions. As we will see later, other electron transport chains can use other compounds besides molecular oxygen as a terminal electron acceptor, such as nitrate or nitrite.

Photophosphorylation
Finally, many organisms can convert the energy from sun light to chemical energy (ATP) which is half of the reactions used in photosynthesis; the other half being the conversion of CO2 to carbohydrates. In photosynthetic organisms, ATP and reducing power (production of NADPH+H+)is generated via a modified electron transport chain where the initial high energy electon(s) is derived from light energy instead of chemical energy from NADH+H+.

While we will discuss this in detail later, in essence oxidative phosphorylation works by taking a reduced high energy compound, such as NADH or FADH2, and transferring electrons in a series of red/ox reactions to other compounds. The energy that is release is then "captured" by the translocation of protons (H+) across the membrane. The net result is an "energize membrane" that separates charge, negative on the inside and positive on the outside (the electrical potential) and the separation of protons (chemical potential), protons concentration greater on the outside then inside of the membrane. The overall energized state of the membrane is then similar to a charged capacitor ready to do work. ATP generation occurs by the F0F1 membrane bound ATPase that translocates protons across the membrane discharging the chemical and electrical potentials with the simultaneous synthesis of ATP. For every three protons (H+) translocated, the F0F1ATPase can synthesize 1 ATP.

Section Summary

ATP functions as the energy currency for cells. It allows the cell to store energy briefly and transport it within the cell to support endergonic chemical reactions. The structure of ATP is that of an RNA nucleotide with three phosphates attached. As ATP is used for energy, a phosphate group or two are detached, and either ADP or AMP is produced. Energy derived from glucose catabolism is used to convert ADP into ATP. When ATP is used in a reaction, the third phosphate is temporarily attached to a substrate in a process called phosphorylation. The two processes of ATP regeneration that are used in conjunction with glucose catabolism are substrate-level phosphorylation and oxidative phosphorylation through the process of chemiosmosis.

Additional Links

Chemwiki Links

- ATP

Insert paragraph text here.

Review Questions

Exercise:

Problem: The energy currency used by cells is _____.

a. ATP
b. ADP
c. AMP
d. adenosine

Solution:

A

Exercise:

Problem:A reducing chemical reaction _____.

a. reduces the compound to a simpler form
b. adds an electron to the substrate
c. removes a hydrogen atom from the substrate
d. is a catabolic reaction

Solution:

B

Free Response

Exercise:

Problem:

Why is it beneficial for cells to use ATP rather than energy directly from the bonds of carbohydrates? What are the greatest drawbacks to harnessing energy directly from the bonds of several different compounds?

Solution:

ATP provides the cell with a way to handle energy in an efficient manner. The molecule can be charged, stored, and used as needed. Moreover, the energy from hydrolyzing ATP is delivered as a consistent amount. Harvesting energy from the bonds of several different compounds would result in energy deliveries of different quantities.

Glossary

chemiosmosis
: process in which there is a production of adenosine triphosphate (ATP) in cellular metabolism by the involvement of a proton gradient across a membrane

dephosphorylation
: removal of a phosphate group from a molecule

oxidative phosphorylation
: production of ATP using the process of chemiosmosis and oxygen

phosphorylation
: addition of a high-energy phosphate to a compound, usually a metabolic intermediate, a protein, or ADP

redox reaction
: chemical reaction that consists of the coupling of an oxidation reaction and a reduction reaction

substrate-level phosphorylation
: production of ATP from ADP using the excess energy from a chemical reaction and a phosphate group from a reactant

Free Energy
By the end of this section, you will be able to:

- Define Gibbs free energy, and describe its relation to spontaneity
- Calculate free energy change for a process using free energies of formation for its reactants and products
- Calculate free energy change for a process using enthalpies of formation and the entropies for its reactants and products
- Explain how temperature affects the spontaneity of some processes
- Relate standard free energy changes to equilibrium constants

One of the challenges of using the second law of thermodynamics to determine if a process is spontaneous is that we must determine the entropy change for the system *and* the entropy change for the surroundings. An alternative approach involving a new thermodynamic property defined in terms of system properties only was introduced in the late nineteenth century by American mathematician Josiah Willard Gibbs. This new property is called the **Gibbs free energy change (*G*)** (or simply the *free energy*), and it is defined in terms of a system's enthalpy and entropy as the following:
Equation:

$$G = H - TS$$

Free energy is a state function, and at constant temperature and pressure, the **standard free energy change (Δ*G*°)** may be expressed as the following:
Equation:

$$\Delta G = \Delta H - T\Delta S$$

(For simplicity's sake, the subscript "sys" will be omitted henceforth.)

We can understand the relationship between this system property and the spontaneity of a process by recalling the previously derived second law expression:
Equation:

$$\Delta S_{\text{univ}} = \Delta S + \frac{q_{\text{surr}}}{T}$$

The first law requires that $q_{\text{surr}} = -q_{\text{sys}}$, and at constant pressure $q_{\text{sys}} = \Delta H$, and so this expression may be rewritten as the following:
Equation:

$$\Delta S_{\text{univ}} = \Delta S - \frac{\Delta H}{T}$$

ΔH is the enthalpy change *of the system*. Multiplying both sides of this equation by $-T$, and rearranging yields the following:
Equation:

$$-T\Delta S_{\text{univ}} = \Delta H - T\Delta S$$

Comparing this equation to the previous one for free energy change shows the following relation:
Equation:

$$\Delta G = -T\Delta S_{\text{univ}}$$

The free energy change is therefore a reliable indicator of the spontaneity of a process, being directly related to the previously identified spontaneity indicator, ΔS_{univ}. [link] summarizes the relation between the spontaneity of a process and the arithmetic signs of these indicators.

Relation between Process Spontaneity and Signs of Thermodynamic Properties		
$\Delta S_{univ} > 0$	$\Delta G < 0$	spontaneous
$\Delta S_{univ} < 0$	$\Delta G > 0$	nonspontaneous
$\Delta S_{univ} = 0$	$\Delta G = 0$	reversible (at equilibrium)

Calculating Free Energy Change

Free energy is a state function, so its value depends only on the conditions of the initial and final states of the system. A convenient and common approach to the calculation of free energy changes for physical and chemical reactions is by use of widely available compilations of standard state thermodynamic data. One method involves the use of standard enthalpies and entropies to compute standard free energy changes according to the following relation as demonstrated in [link].

Equation:

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$$

Example:

Evaluation of Δ*G*° from Δ*H*° and Δ*S*°

Use standard enthalpy and entropy data from Appendix G to calculate the standard free energy change for the vaporization of water at room temperature (298 K). What does the computed value for Δ*G*° say about the spontaneity of this process?

Solution

The process of interest is the following:

Equation:

$$\mathrm{H_2O}(l) \longrightarrow \mathrm{H_2O}(g)$$

The standard change in free energy may be calculated using the following equation:

Equation:

$$\Delta \overset{\circ}{G}_{298} = \Delta H^\circ - T\Delta S^\circ$$

From Appendix G, here is the data:

Substance	$\Delta \overset{\circ}{H}_f\ (\mathrm{kJ/mol})$	$\overset{\circ}{S}_{298}(\mathrm{J/K{\cdot}mol})$

Substance	$\Delta H_{\mathrm{f}}^{\circ}\,(\mathrm{kJ/mol})$	$S_{298}^{\circ}(\mathrm{J/K{\cdot}mol})$
$H_2O(l)$	−286.83	70.0
$H_2O(g)$	−241.82	188.8

Combining at 298 K:
Equation:

$$\Delta H^{\circ} = \Delta H_{298}^{\circ} = \Delta H_{\mathrm{f}}^{\circ}\,(\mathrm{H_2O}(g)) - \Delta H_{\mathrm{f}}^{\circ}\,(\mathrm{H_2O}(l))$$
$$= [-241.82\ \mathrm{kJ} - (-286.83)]\ \mathrm{kJ/mol} = 44.01\ \mathrm{kJ/mol}$$

Equation:

$$\Delta S^{\circ} = \Delta S_{298}^{\circ} = S_{298}^{\circ}\,(\mathrm{H_2O}(g)) - S_{298}^{\circ}\,(\mathrm{H_2O}(l))$$
$$= 188.8\ \mathrm{J/mol{\cdot}K} - 70.0\ \mathrm{J/K} = 118.8\ \mathrm{J/mol{\cdot}K}$$

Equation:

$$\Delta G^{\circ} = \Delta H^{\circ} - T\Delta S^{\circ}$$

Converting everything into kJ and combining at 298 K:
Equation:

$$\Delta G_{298}^{\circ} = \Delta H^{\circ} - T\Delta S^{\circ}$$
$$= 44.01\ \mathrm{kJ/mol} - (298\ \mathrm{K} \times 118.8\ \mathrm{J/mol{\cdot}K}) \times \tfrac{1\ \mathrm{kJ}}{1000\ \mathrm{J}}$$

Equation:

$$44.01\ \mathrm{kJ/mol} - 35.4\ \mathrm{kJ/mol} = 9.6\ \mathrm{kJ/mol}$$

At 298 K (25 °C) $\Delta G_{298}^{\circ} > 0$, and so boiling is nonspontaneous (*not* spontaneous).
Check Your Learning
Use standard enthalpy and entropy data from Appendix G to calculate the standard free energy change for the reaction shown here (298 K). What does the computed value for Δ*G*° say about the spontaneity of this process?
Equation:

$$\mathrm{C_2H_6}(g) \longrightarrow \mathrm{H_2}(g) + \mathrm{C_2H_4}(g)$$

Note:
Answer:
$\Delta G_{298}^{\circ} = 102.0\ \mathrm{kJ/mol}$; the reaction is nonspontaneous (*not* spontaneous) at 25 °C.

The standard free energy change for a reaction may also be calculated from **standard free energy of formation** $(\Delta G_{\mathrm{f}}^{\circ})$, values of the reactants and products involved in the reaction. The standard free energy of formation is the

free energy change that accompanies the formation of one mole of a substance from its elements in their standard states. Similar to the standard enthalpy of formation, $\Delta G_{\mathrm{f}}^{\circ}$ is by definition zero for elemental substances under standard state conditions. The approach used to calculate ΔG° for a reaction from $\Delta G_{\mathrm{f}}^{\circ}$ values is the same as that demonstrated previously for enthalpy and entropy changes. For the reaction

Equation:

$$m\mathrm{A} + n\mathrm{B} \longrightarrow x\mathrm{C} + y\mathrm{D},$$

the standard free energy change at room temperature may be calculated as

Equation:

$$\begin{aligned}\Delta G_{298}^{\circ} = \Delta G^{\circ} &= \sum \nu \Delta G_{298}^{\circ}(\text{products}) - \sum \nu \Delta G_{298}^{\circ}(\text{reactants}) \\ &= \left[x\Delta G_{\mathrm{f}}^{\circ}(\mathrm{C}) + y\Delta G_{\mathrm{f}}^{\circ}(\mathrm{D})\right] - \left[m\Delta G_{\mathrm{f}}^{\circ}(\mathrm{A}) + n\Delta G_{\mathrm{f}}^{\circ}(\mathrm{B})\right].\end{aligned}$$

Example:

Calculation of ΔG_{298}°

Consider the decomposition of yellow mercury(II) oxide.

Equation:

$$\mathrm{HgO}(s,\ \text{yellow}) \longrightarrow \mathrm{Hg}(l) + \frac{1}{2}\mathrm{O}_2(g)$$

Calculate the standard free energy change at room temperature, ΔG_{298}°, using (a) standard free energies of formation and (b) standard enthalpies of formation and standard entropies. Do the results indicate the reaction to be spontaneous or nonspontaneous under standard conditions?

Solution

The required data are available in Appendix G and are shown here.

Compound	$\Delta G_{\mathrm{f}}^{\circ}$ (kJ/mol)	$\Delta H_{\mathrm{f}}^{\circ}$ (kJ/mol)	S_{298}° (J/K·mol)
HgO (*s*, yellow)	−58.43	−90.46	71.13
Hg(*l*)	0	0	75.9
O_2(*g*)	0	0	205.2

(a) Using free energies of formation:

Equation:

$$\Delta G_{298}^{\circ} = \sum \nu G S_{\mathrm{f}}^{\circ}(\text{products}) - \sum \nu \Delta G_{\mathrm{f}}^{\circ}(\text{reactants})$$

Equation:

$$= \left[1\Delta G_{\mathrm{f}}^{\circ}\mathrm{Hg}(l) + \frac{1}{2}\Delta G_{\mathrm{f}}^{\circ}\mathrm{O_2}(g)\right] - 1\Delta G_{\mathrm{f}}^{\circ}\mathrm{HgO}(s,\text{ yellow})$$

Equation:

$$= \left[1\text{ mol}(0\text{ kJ/mol}) + \frac{1}{2}\text{ mol}(0\text{ kJ/mol})\right] - 1\text{ mol}(-58.43\text{ kJ/mol}) = 58.43\text{ kJ/mol}$$

(b) Using enthalpies and entropies of formation:

Equation:

$$\Delta H_{298}^{\circ} = \sum \nu \Delta H_{\mathrm{f}}^{\circ}\,(\text{products}) - \sum \nu \Delta H_{\mathrm{f}}^{\circ}\,(\text{reactants})$$

Equation:

$$= \left[1\Delta H_{\mathrm{f}}^{\circ}\mathrm{Hg}(l) + \frac{1}{2}\Delta H_{\mathrm{f}}^{\circ}\mathrm{O_2}(g)\right] - 1\Delta H_{\mathrm{f}}^{\circ}\mathrm{HgO}(s,\text{ yellow})$$

Equation:

$$= \left[1\text{ mol}(0\text{ kJ/mol}) + \frac{1}{2}\text{ mol}(0\text{ kJ/mol})\right] - 1\text{ mol}(-90.46\text{ kJ/mol}) = 90.46\text{ kJ/mol}$$

Equation:

$$\Delta S_{298}^{\circ} = \sum \mathrm{v}\Delta S_{298}^{\circ}(\text{products}) - \sum \mathrm{v}\Delta S_{298}^{\circ}(\text{reactants})$$

Equation:

$$= \left[1\Delta S_{298}^{\circ}\mathrm{Hg}(l) + \frac{1}{2}\Delta S_{298}^{\circ}\mathrm{O_2}(g)\right] - 1\Delta S_{298}^{\circ}\mathrm{HgO}(s,\text{ yellow})$$

Equation:

$$= \left[1\text{ mol}\,(75.9\text{ J/mol K}) + \frac{1}{2}\text{ mol}(205.2\text{ J/mol K})\right] - 1\text{ mol}(71.13\text{ J/mol K}) = 107.4\text{ J/mol K}$$

Equation:

$$\Delta G^{\circ} = \Delta H^{\circ} - T\Delta S^{\circ} = 90.46\text{ kJ} - 298.15\text{ K} \times 107.4\text{ J/K·mol} \times \frac{1\text{ kJ}}{1000\text{ J}}$$

Equation:

$$\Delta G^{\circ} = (90.46 - 32.01)\text{ kJ/mol} = 58.45\text{ kJ/mol}$$

Both ways to calculate the standard free energy change at 25 °C give the same numerical value (to three significant figures), and both predict that the process is nonspontaneous (*not* spontaneous) at room temperature.

Check Your Learning

Calculate Δ*G*° using (a) free energies of formation and (b) enthalpies of formation and entropies (Appendix G). Do the results indicate the reaction to be spontaneous or nonspontaneous at 25 °C?

Equation:

$$\mathrm{C_2H_4}(g) \longrightarrow \mathrm{H_2}(g) + \mathrm{C_2H_2}(g)$$

Note:
Answer:
(a) 140.8 kJ/mol, nonspontaneous
(b) 141.5 kJ/mol, nonspontaneous

Temperature Dependence of Spontaneity

As was previously demonstrated in this chapter's section on entropy, the spontaneity of a process may depend upon the temperature of the system. Phase transitions, for example, will proceed spontaneously in one direction or the other depending upon the temperature of the substance in question. Likewise, some chemical reactions can also exhibit temperature dependent spontaneities. To illustrate this concept, the equation relating free energy change to the enthalpy and entropy changes for the process is considered:
Equation:

$$\Delta G = \Delta H - T\Delta S$$

The spontaneity of a process, as reflected in the arithmetic sign of its free energy change, is then determined by the signs of the enthalpy and entropy changes and, in some cases, the absolute temperature. Since T is the absolute (kelvin) temperature, it can only have positive values. Four possibilities therefore exist with regard to the signs of the enthalpy and entropy changes:

1. **Both ΔH and ΔS are positive.** This condition describes an endothermic process that involves an increase in system entropy. In this case, ΔG will be negative if the magnitude of the $T\Delta S$ term is greater than ΔH. If the $T\Delta S$ term is less than ΔH, the free energy change will be positive. Such a process is *spontaneous at high temperatures and nonspontaneous at low temperatures.*
2. **Both ΔH and ΔS are negative.** This condition describes an exothermic process that involves a decrease in system entropy. In this case, ΔG will be negative if the magnitude of the $T\Delta S$ term is less than ΔH. If the $T\Delta S$ term's magnitude is greater than ΔH, the free energy change will be positive. Such a process is *spontaneous at low temperatures and nonspontaneous at high temperatures.*
3. **ΔH is positive and ΔS is negative.** This condition describes an endothermic process that involves a decrease in system entropy. In this case, ΔG will be positive regardless of the temperature. Such a process is *nonspontaneous at all temperatures.*
4. **ΔH is negative and ΔS is positive.** This condition describes an exothermic process that involves an increase in system entropy. In this case, ΔG will be negative regardless of the temperature. Such a process is *spontaneous at all temperatures.*

These four scenarios are summarized in [link].

Summary of the Four Scenarios for Enthalpy and Entropy Changes

	ΔH > 0 (endothermic)	**ΔH < 0 (exothermic)**
ΔS > 0 (increase in entropy)	ΔG < 0 at high temperature ΔG > 0 at low temperature Process is spontaneous at high temperature	ΔG < 0 at any temperature Process is spontaneous at any temperature
ΔS < 0 (decrease in entropy)	ΔG > 0 at any temperature Process is nonspontaneous at any temperature	ΔG < 0 at low temperature ΔG > 0 at high temperature Process is spontaneous at low temperature

There are four possibilities regarding the signs of enthalpy and entropy changes.

Example:
Predicting the Temperature Dependence of Spontaneity
The incomplete combustion of carbon is described by the following equation:
Equation:

$$2\mathrm{C}(s) + \mathrm{O}_2(g) \longrightarrow 2\mathrm{CO}(g)$$

How does the spontaneity of this process depend upon temperature?
Solution
Combustion processes are exothermic ($\Delta H < 0$). This particular reaction involves an increase in entropy due to the accompanying increase in the amount of gaseous species (net gain of one mole of gas, $\Delta S > 0$). The reaction is therefore spontaneous ($\Delta G < 0$) at all temperatures.
Check Your Learning
Popular chemical hand warmers generate heat by the air-oxidation of iron:
Equation:

$$4\mathrm{Fe}(s) + 3\mathrm{O}_2(g) \longrightarrow 2\mathrm{Fe}_2\mathrm{O}_3(s)$$

How does the spontaneity of this process depend upon temperature?

Note:
Answer:
ΔH and ΔS are negative; the reaction is spontaneous at low temperatures.

When considering the conclusions drawn regarding the temperature dependence of spontaneity, it is important to keep in mind what the terms “high” and “low” mean. Since these terms are adjectives, the temperatures in question are deemed high or low relative to some reference temperature. A process that is nonspontaneous at one temperature but spontaneous at another will necessarily undergo a change in “spontaneity” (as reflected by its ΔG)

as temperature varies. This is clearly illustrated by a graphical presentation of the free energy change equation, in which Δ*G* is plotted on the *y* axis versus *T* on the *x* axis:
Equation:

$$\Delta G = \Delta H - T\Delta S$$

Equation:

$$y = b + mx$$

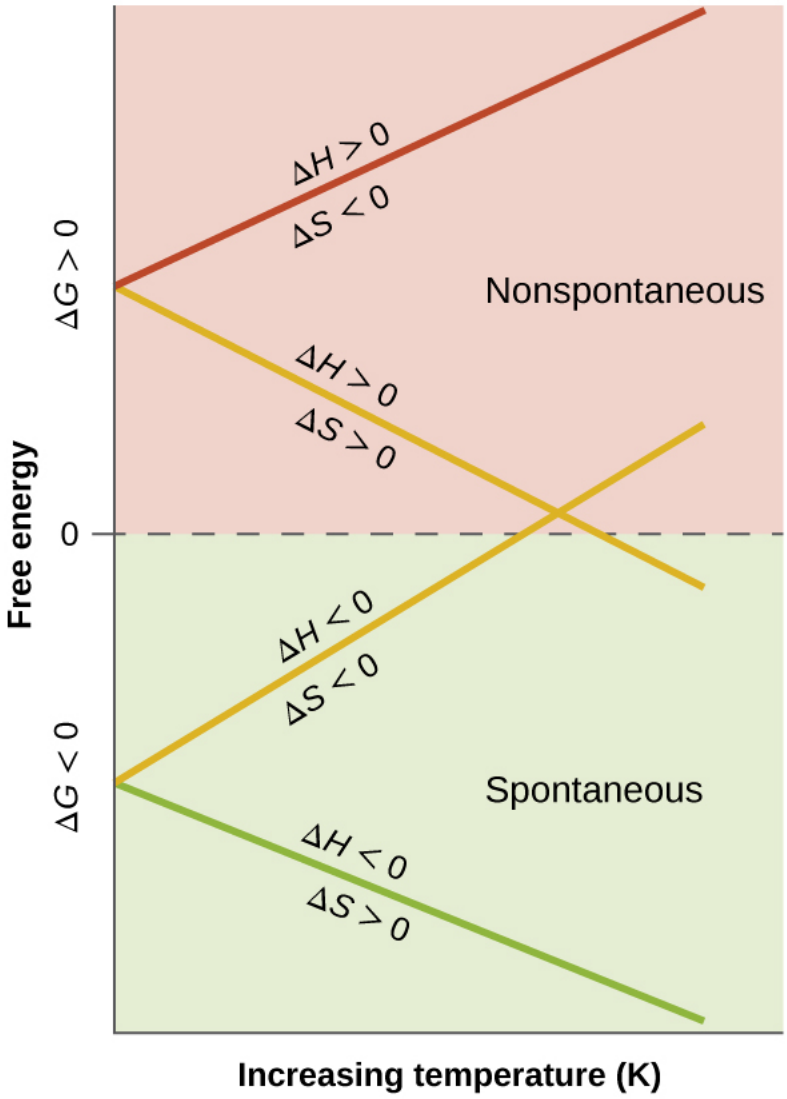
Such a plot is shown in [link]. A process whose enthalpy and entropy changes are of the same arithmetic sign will exhibit a temperature-dependent spontaneity as depicted by the two yellow lines in the plot. Each line crosses from one spontaneity domain (positive or negative Δ*G*) to the other at a temperature that is characteristic of the process in question. This temperature is represented by the *x*-intercept of the line, that is, the value of *T* for which Δ*G* is zero:
Equation:

$$\Delta G = 0 = \Delta H - T\Delta S$$

Equation:

$$T = \frac{\Delta H}{\Delta S}$$

And so, saying a process is spontaneous at “high” or “low” temperatures means the temperature is above or below, respectively, that temperature at which Δ*G* for the process is zero. As noted earlier, the condition of ΔG = 0 describes a system at equilibrium.



These plots show the variation in ΔG with temperature for the four possible combinations of arithmetic sign for ΔH and ΔS.

Example:
Equilibrium Temperature for a Phase Transition
As defined in the chapter on liquids and solids, the boiling point of a liquid is the temperature at which its liquid and gaseous phases are in equilibrium (that is, when vaporization and condensation occur at equal rates). Use the information in Appendix G to estimate the boiling point of water.
Solution
The process of interest is the following phase change:
Equation:

$$\mathrm{H_2O}(l) \longrightarrow \mathrm{H_2O}(g)$$

When this process is at equilibrium, $\Delta G = 0$, so the following is true:
Equation:

$$0 = \Delta H^\circ - T\Delta S^\circ \qquad \text{or} \qquad T = \frac{\Delta H^\circ}{\Delta S^\circ}$$

Using the standard thermodynamic data from Appendix G,
Equation:

$$\begin{aligned}\Delta H^\circ &= \Delta H_\text{f}^\circ\,(\text{H}_2\text{O}(g)) - \Delta H_\text{f}^\circ\,(\text{H}_2\text{O}(l)) \\ &= -241.82\ \text{kJ/mol} - (-286.83\ \text{kJ/mol}) = 44.01\ \text{kJ/mol}\end{aligned}$$

Equation:

$$\begin{aligned}\Delta S^\circ &= \Delta S_{298}^\circ\,(\text{H}_2\text{O}(g)) - \Delta S_{298}^\circ\,(\text{H}_2\text{O}(l)) \\ &= 188.8\ \text{J/K·mol} - 70.0\ \text{J/K·mol} = 118.8\ \text{J/K·mol}\end{aligned}$$

Equation:

$$T = \frac{\Delta H^\circ}{\Delta S^\circ} = \frac{44.01 \times 10^3\ \text{J/mol}}{118.8\ \text{J/K·mol}} = 370.5\ \text{K} = 97.3\ ^\circ\text{C}$$

The accepted value for water's normal boiling point is 373.2 K (100.0 °C), and so this calculation is in reasonable agreement. Note that the values for enthalpy and entropy changes data used were derived from standard data at 298 K (Appendix G). If desired, you could obtain more accurate results by using enthalpy and entropy changes determined at (or at least closer to) the actual boiling point.
Check Your Learning
Use the information in Appendix G to estimate the boiling point of CS_2.

Note:
Answer:
313 K (accepted value 319 K)

Free Energy and Equilibrium

The free energy change for a process may be viewed as a measure of its driving force. A negative value for ΔG represents a finite driving force for the process in the forward direction, while a positive value represents a driving force for the process in the reverse direction. When ΔG is zero, the forward and reverse driving forces are equal, and so the process occurs in both directions at the same rate (the system is at equilibrium).

In the chapter on equilibrium the *reaction quotient*, *Q*, was introduced as a convenient measure of the status of an equilibrium system. Recall that *Q* is the numerical value of the mass action expression for the system, and that you may use its value to identify the direction in which a reaction will proceed in order to achieve equilibrium. When *Q* is lesser than the equilibrium constant, *K*, the reaction will proceed in the forward direction until equilibrium is reached and $Q = K$. Conversely, if $Q > K$, the process will proceed in the reverse direction until equilibrium is achieved.

The free energy change for a process taking place with reactants and products present under nonstandard conditions, ΔG, is related to the standard free energy change, $\Delta G°$, according to this equation:

Equation:

$$\Delta G = \Delta G^\circ + RT \ln Q$$

R is the gas constant (8.314 J/K mol), *T* is the kelvin or absolute temperature, and *Q* is the reaction quotient. We may use this equation to predict the spontaneity for a process under any given set of conditions as illustrated in [link].

Example:
Calculating Δ*G* under Nonstandard Conditions
What is the free energy change for the process shown here under the specified conditions?
T = 25 °C, $P_{\mathrm{N_2}} = 0.870$ atm, $P_{\mathrm{H_2}} = 0.250$ atm, and $P_{\mathrm{NH_3}} = 12.9$ atm
Equation:

$$2\mathrm{NH_3}(g) \longrightarrow 3\mathrm{H_2}(g) + \mathrm{N_2}(g) \qquad \Delta G^\circ = 33.0\ \mathrm{kJ/mol}$$

Solution
The equation relating free energy change to standard free energy change and reaction quotient may be used directly:
Equation:

$$\Delta G = \Delta G^\circ + RT \ln Q = 33.0\ \frac{\mathrm{kJ}}{\mathrm{mol}} + \left(8.314\ \frac{\mathrm{J}}{\mathrm{mol\ K}} \times 298\ \mathrm{K} \times \ln \frac{\left(0.250^3\right) \times 0.870}{12.9^2}\right) = 9680\ \frac{\mathrm{J}}{\mathrm{mol}}$$

Since the computed value for Δ*G* is positive, the reaction is nonspontaneous under these conditions.
Check Your Learning
Calculate the free energy change for this same reaction at 875 °C in a 5.00 L mixture containing 0.100 mol of each gas. Is the reaction spontaneous under these conditions?

Note:
Answer:
Δ*G* = −47 kJ; yes

For a system at equilibrium, *Q* = *K* and Δ*G* = 0, and the previous equation may be written as
Equation:

$$0 = \Delta G^\circ + RT \ln K \qquad \text{(at equilibrium)}$$

Equation:

$$\Delta G^\circ = -RT \ln K \qquad \text{or} \qquad K = e^{-\frac{\Delta G^\circ}{RT}}$$

This form of the equation provides a useful link between these two essential thermodynamic properties, and it can be used to derive equilibrium constants from standard free energy changes and vice versa. The relations between standard free energy changes and equilibrium constants are summarized in [link].

Relations between Standard Free Energy Changes and Equilibrium Constants		
K	**Δ*G*°**	**Comments**
> 1	< 0	Products are more abundant at equilibrium.
< 1	> 0	Reactants are more abundant at equilibrium.
= 1	= 0	Reactants and products are equally abundant at equilibrium.

Example:
Calculating an Equilibrium Constant using Standard Free Energy Change
Given that the standard free energies of formation of $Ag^+(aq)$, $Cl^-(aq)$, and $AgCl(s)$ are 77.1 kJ/mol, −131.2 kJ/mol, and −109.8 kJ/mol, respectively, calculate the solubility product, K_{sp}, for AgCl.
Solution
The reaction of interest is the following:
Equation:

$$\mathrm{AgCl}(s) \rightleftharpoons \mathrm{Ag^+}(aq) + \mathrm{Cl^-}(aq) \qquad\qquad K_{\mathrm{sp}} = [\mathrm{Ag^+}][\mathrm{Cl^-}]$$

The standard free energy change for this reaction is first computed using standard free energies of formation for its reactants and products:
Equation:

$$\begin{aligned}\Delta G^\circ = \Delta \overset{\circ}{G}_{298} &= \left[\Delta \overset{\circ}{G}_{\mathrm{f}}\left(\mathrm{Ag^+}(aq)\right) + \Delta \overset{\circ}{G}_{\mathrm{f}}\left(\mathrm{Cl^-}(aq)\right)\right] - \left[\Delta \overset{\circ}{G}_{\mathrm{f}}\left(\mathrm{AgCl}(s)\right)\right] \\ &= [77.1\ \mathrm{kJ/mol} - 131.2\ \mathrm{kJ/mol}] - [-109.8\ \mathrm{kJ/mol}] = 55.7\ \mathrm{kJ/mol}\end{aligned}$$

The equilibrium constant for the reaction may then be derived from its standard free energy change:
Equation:

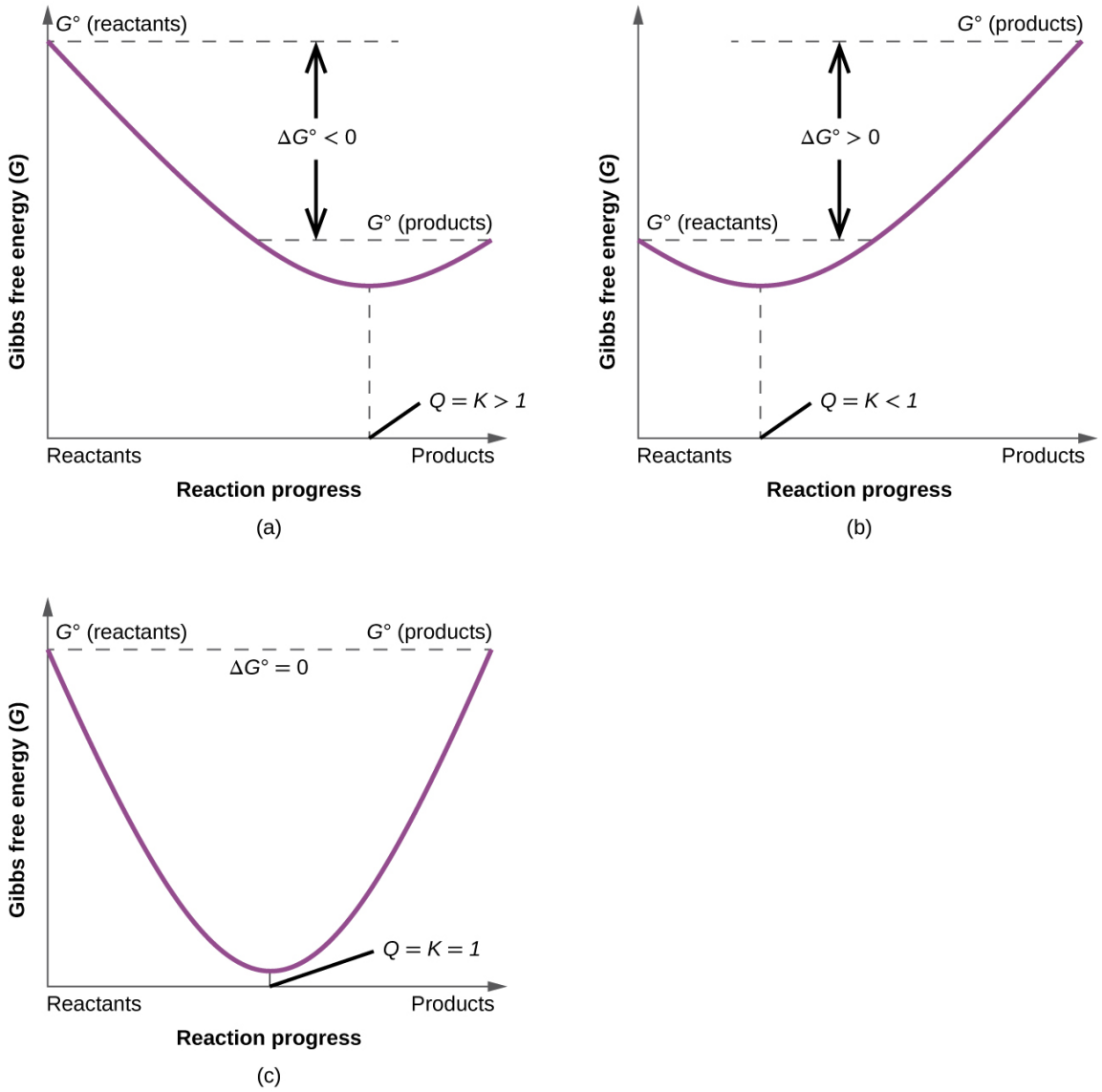
$$K_{\mathrm{sp}} = e^{-\frac{\Delta G^\circ}{RT}} = \exp\left(-\frac{\Delta G^\circ}{RT}\right) = \exp\left(-\frac{55.7 \times 10^3\ \mathrm{J/mol}}{8.314\ \mathrm{J/mol{\cdot}K} \times 298.15\ \mathrm{K}}\right) = \exp\left(-22.470\right) = e^{-22.470} = 1.$$

This result is in reasonable agreement with the value provided in Appendix J.
Check Your Learning
Use the thermodynamic data provided in Appendix G to calculate the equilibrium constant for the dissociation of dinitrogen tetroxide at 25 °C.
Equation:

$$2\mathrm{NO_2}(g) \rightleftharpoons \mathrm{N_2O_4}(g)$$

Note:
Answer:
K = 6.9

To further illustrate the relation between these two essential thermodynamic concepts, consider the observation that reactions spontaneously proceed in a direction that ultimately establishes equilibrium. As may be shown by plotting the free energy change versus the extent of the reaction (for example, as reflected in the value of Q), equilibrium is established when the system's free energy is minimized ([link]). If a system consists of reactants and products in nonequilibrium amounts ($Q \neq K$), the reaction will proceed spontaneously in the direction necessary to establish equilibrium.



These plots show the free energy versus reaction progress for systems whose standard free energy changes are (a) negative, (b) positive, and (c) zero. Nonequilibrium systems will proceed spontaneously in whatever direction is necessary to minimize free energy and establish equilibrium.

Key Concepts and Summary

Gibbs free energy (G) is a state function defined with regard to system quantities only and may be used to predict the spontaneity of a process. A negative value for ΔG indicates a spontaneous process; a positive ΔG indicates a nonspontaneous process; and a ΔG of zero indicates that the system is at equilibrium. A number of approaches to the computation of free energy changes are possible.

Key Equations

- $\Delta G = \Delta H - T\Delta S$
- $\Delta G = \Delta G° + RT \ln Q$
- $\Delta G° = -RT \ln K$

Chemistry End of Chapter Exercises

Exercise:

Problem: What is the difference between ΔG, $\Delta G°$, and $\Delta \overset{\circ}{G}_{298}$ for a chemical change?

Exercise:

Problem:

A reaction has $\Delta \overset{\circ}{H}_{298}$ = 100 kJ/mol and $\Delta \overset{\circ}{S}_{298} = 250\ \text{J/mol·K}$. Is the reaction spontaneous at room temperature? If not, under what temperature conditions will it become spontaneous?

Solution:

The reaction is nonspontaneous at room temperature.
Above 400 K, ΔG will become negative, and the reaction will become spontaneous.

Exercise:

Problem:

Explain what happens as a reaction starts with $\Delta G < 0$ (negative) and reaches the point where $\Delta G = 0$.

Exercise:

Problem:

Use the standard free energy of formation data in Appendix G to determine the free energy change for each of the following reactions, which are run under standard state conditions and 25 °C. Identify each as either spontaneous or nonspontaneous at these conditions.

(a) $\text{MnO}_2(s) \longrightarrow \text{Mn}(s) + \text{O}_2(g)$

(b) $\text{H}_2(g) + \text{Br}_2(l) \longrightarrow 2\text{HBr}(g)$

(c) $\text{Cu}(s) + \text{S}(g) \longrightarrow \text{CuS}(s)$

(d) $2\text{LiOH}(s) + \text{CO}_2(g) \longrightarrow \text{Li}_2\text{CO}_3(s) + \text{H}_2\text{O}(g)$

(e) $\text{CH}_4(g) + \text{O}_2(g) \longrightarrow \text{C}(s,\ \text{graphite}) + 2\text{H}_2\text{O}(g)$

(f) $\text{CS}_2(g) + 3\text{Cl}_2(g) \longrightarrow \text{CCl}_4(g) + \text{S}_2\text{Cl}_2(g)$

Solution:

(a) 465.1 kJ nonspontaneous; (b) −106.86 kJ spontaneous; (c) −53.6 kJ spontaneous; (d) −83.4 kJ spontaneous; (e) −406.7 kJ spontaneous; (f) −30.0 kJ spontaneous

Exercise:

Problem:

Use the standard free energy data in Appendix G to determine the free energy change for each of the following reactions, which are run under standard state conditions and 25 °C. Identify each as either spontaneous or nonspontaneous at these conditions.

(a) $\mathrm{C}(s,\text{ graphite}) + \mathrm{O_2}(g) \longrightarrow \mathrm{CO_2}(g)$

(b) $\mathrm{O_2}(g) + \mathrm{N_2}(g) \longrightarrow 2\mathrm{NO}(g)$

(c) $2\mathrm{Cu}(s) + \mathrm{S}(g) \longrightarrow \mathrm{Cu_2S}(s)$

(d) $\mathrm{CaO}(s) + \mathrm{H_2O}(l) \longrightarrow \mathrm{Ca(OH)_2}(s)$

(e) $\mathrm{Fe_2O_3}(s) + 3\mathrm{CO}(g) \longrightarrow 2\mathrm{Fe}(s) + 3\mathrm{CO_2}(g)$

(f) $\mathrm{CaSO_4{\cdot}2H_2O}(s) \longrightarrow \mathrm{CaSO_4}(s) + 2\mathrm{H_2O}(g)$

Exercise:

Given:

$\mathrm{P_4}(s) + 5\mathrm{O_2}(g) \longrightarrow \mathrm{P_4O_{10}}(s) \qquad \Delta G^\circ_{298} = -2697.0\text{ kJ/mol}$

$2\mathrm{H_2}(g) + \mathrm{O_2}(g) \longrightarrow 2\mathrm{H_2O}(g) \qquad \Delta G^\circ_{298} = -457.18\text{ kJ/mol}$

Problem: $6\mathrm{H_2O}(g) + \mathrm{P_4O_{10}}(s) \longrightarrow 4\mathrm{H_3PO_4}(l) \qquad \Delta G^\circ_{298} = -428.66\text{ kJ/mol}$

(a) Determine the standard free energy of formation, $\Delta G^\circ_\mathrm{f}$, for phosphoric acid.

(b) How does your calculated result compare to the value in Appendix G? Explain.

Solution:

(a) The standard free energy of formation is –1124.3 kJ/mol. (b) The calculation agrees with the value in Appendix G because free energy is a state function (just like the enthalpy and entropy), so its change depends only on the initial and final states, not the path between them.

Exercise:

Problem:

Is the formation of ozone ($O_3(g)$) from oxygen ($O_2(g)$) spontaneous at room temperature under standard state conditions?

Exercise:

Consider the decomposition of red mercury(II) oxide under standard state conditions.

Problem: $2\mathrm{HgO}(s,\text{ red}) \longrightarrow 2\mathrm{Hg}(l) + \mathrm{O_2}(g)$

(a) Is the decomposition spontaneous under standard state conditions?

(b) Above what temperature does the reaction become spontaneous?

Solution:

(a) The reaction is nonspontaneous; (b) Above 566 °C the process is spontaneous.

Exercise:

Problem:

Among other things, an ideal fuel for the control thrusters of a space vehicle should decompose in a spontaneous exothermic reaction when exposed to the appropriate catalyst. Evaluate the following substances under standard state conditions as suitable candidates for fuels.

(a) Ammonia: $2\mathrm{NH_3}(g) \longrightarrow \mathrm{N_2}(g) + 3\mathrm{H_2}(g)$

(b) Diborane: $\mathrm{B_2H_6}(g) \longrightarrow 2\mathrm{B}(g) + 3\mathrm{H_2}(g)$

(c) Hydrazine: $\mathrm{N_2H_4}(g) \longrightarrow \mathrm{N_2}(g) + 2\mathrm{H_2}(g)$

(d) Hydrogen peroxide: $\mathrm{H_2O_2}(l) \longrightarrow \mathrm{H_2O}(g) + \frac{1}{2}\mathrm{O_2}(g)$

Exercise:

Problem:

Calculate Δ*G*° for each of the following reactions from the equilibrium constant at the temperature given.

(a) $\mathrm{N_2}(g) + \mathrm{O_2}(g) \longrightarrow 2\mathrm{NO}(g) \qquad \mathrm{T} = 2000\ ^\circ\mathrm{C} \qquad K_p = 4.1 \times 10^{-4}$

(b) $\mathrm{H_2}(g) + \mathrm{I_2}(g) \longrightarrow 2\mathrm{HI}(g) \qquad \mathrm{T} = 400\ ^\circ\mathrm{C} \qquad K_p = 50.0$

(c) $\mathrm{CO_2}(g) + \mathrm{H_2}(g) \longrightarrow \mathrm{CO}(g) + \mathrm{H_2O}(g) \qquad \mathrm{T} = 980\ ^\circ\mathrm{C} \qquad K_p = 1.67$

(d) $\mathrm{CaCO_3}(s) \longrightarrow \mathrm{CaO}(s) + \mathrm{CO_2}(g) \qquad \mathrm{T} = 900\ ^\circ\mathrm{C} \qquad K_p = 1.04$

(e) $\mathrm{HF}(aq) + \mathrm{H_2O}(l) \longrightarrow \mathrm{H_3O^+}(aq) + \mathrm{F^-}(aq) \qquad \mathrm{T} = 25\ ^\circ\mathrm{C} \qquad K_p = 7.2 \times 10^{-4}$

(f) $\mathrm{AgBr}(s) \longrightarrow \mathrm{Ag^+}(aq) + \mathrm{Br^-}(aq) \qquad \mathrm{T} = 25\ ^\circ\mathrm{C} \qquad K_p = 3.3 \times 10^{-13}$

Solution:

(a) 1.5 × 10^2 kJ; (b) −21.9 kJ; (c) −5.34 kJ; (d) −0.383 kJ; (e) 18 kJ; (f) 71 kJ

Exercise:

Problem:

Calculate Δ*G*° for each of the following reactions from the equilibrium constant at the temperature given.

(a) $\mathrm{Cl_2}(g) + \mathrm{Br_2}(g) \longrightarrow 2\mathrm{BrCl}(g) \qquad \mathrm{T} = 25\ ^\circ\mathrm{C} \qquad K_p = 4.7 \times 10^{-2}$

(b) $2\mathrm{SO_2}(g) + \mathrm{O_2}(g) \rightleftharpoons 2\mathrm{SO_3}(g) \qquad \mathrm{T} = 500\ ^\circ\mathrm{C} \qquad K_p = 48.2$

(c) $\mathrm{H_2O}(l) \rightleftharpoons \mathrm{H_2O}(g) \qquad \mathrm{T} = 60\ ^\circ\mathrm{C} \qquad K_p = 0.196\ \mathrm{atm}$

(d) $\mathrm{CoO}(s) + \mathrm{CO}(g) \rightleftharpoons \mathrm{Co}(s) + \mathrm{CO_2}(g) \qquad \mathrm{T} = 550\ ^\circ\mathrm{C} \qquad K_p = 4.90 \times 10^2$

(e)
$\mathrm{CH_3NH_2}(aq) + \mathrm{H_2O}(l) \longrightarrow \mathrm{CH_3NH_3^+}(aq) + \mathrm{OH^-}(aq) \qquad \mathrm{T} = 25\ ^\circ\mathrm{C} \qquad K_p = 4.4 \times$

(f) $\mathrm{PbI_2}(s) \longrightarrow \mathrm{Pb^{2+}}(aq) + 2\mathrm{I^-}(aq) \qquad \mathrm{T} = 25\ ^\circ\mathrm{C} \qquad K_p = 8.7 \times 10^{-9}$

Exercise:

Problem:

Calculate the equilibrium constant at 25 °C for each of the following reactions from the value of Δ*G*° given.

(a) $\mathrm{O_2}(g) + 2\mathrm{F_2}(g) \longrightarrow 2\mathrm{OF_2}(g) \qquad \Delta G^\circ = -9.2\ \mathrm{kJ}$

(b) $\mathrm{I_2}(s) + \mathrm{Br_2}(l) \longrightarrow 2\mathrm{IBr}(g) \qquad \Delta G^\circ = 7.3\ \mathrm{kJ}$

(c) $2\mathrm{LiOH}(s) + \mathrm{CO_2}(g) \longrightarrow \mathrm{Li_2CO_3}(s) + \mathrm{H_2O}(g) \qquad \Delta G^\circ = -79\ \mathrm{kJ}$

(d) $\mathrm{N_2O_3}(g) \longrightarrow \mathrm{NO}(g) + \mathrm{NO_2}(g) \qquad \Delta G^\circ = -1.6\ \mathrm{kJ}$

(e) $\mathrm{SnCl_4}(l) \longrightarrow \mathrm{SnCl_4}(l) \qquad \Delta G^\circ = 8.0\ \mathrm{kJ}$

Solution:

(a) *K* = 41; (b) *K* = 0.053; (c) *K* = 6.9 × 10^{13}; (d) *K* = 1.9; (e) *K* = 0.04

Exercise:

Problem:

Calculate the equilibrium constant at 25 °C for each of the following reactions from the value of Δ*G*° given.

(a) $\mathrm{I_2}(s) + \mathrm{Cl_2}(g) \longrightarrow 2\mathrm{ICl}(g) \qquad \Delta G^\circ = -10.88\ \mathrm{kJ}$

(b) $\mathrm{H_2}(g) + \mathrm{I_2}(s) \longrightarrow 2\mathrm{HI}(g) \qquad \Delta G^\circ = 3.4\ \mathrm{kJ}$

(c) $\mathrm{CS_2}(g) + 3\mathrm{Cl_2}(g) \longrightarrow \mathrm{CCl_4}(g) + \mathrm{S_2Cl_2}(g) \qquad \Delta G^\circ = -39\ \mathrm{kJ}$

(d) $2\mathrm{SO_2}(g) + \mathrm{O_2}(g) \longrightarrow 2\mathrm{SO_3}(g) \qquad \Delta G^\circ = -141.82\ \mathrm{kJ}$

(e) $\mathrm{CS_2}(g) \longrightarrow \mathrm{CS_2}(l) \qquad \Delta G^\circ = -1.88\ \mathrm{kJ}$

Exercise:

Problem: Calculate the equilibrium constant at the temperature given.

(a) $\mathrm{O_2}(g) + 2\mathrm{F_2}(g) \longrightarrow 2\mathrm{F_2O}(g) \qquad (\mathrm{T} = 100\ ^\circ\mathrm{C})$

(b) $\mathrm{I_2}(s) + \mathrm{Br_2}(l) \longrightarrow 2\mathrm{IBr}(g) \qquad (\mathrm{T} = 0.0\ ^\circ\mathrm{C})$

(c) $2\mathrm{LiOH}(s) + \mathrm{CO_2}(g) \longrightarrow \mathrm{Li_2CO_3}(s) + \mathrm{H_2O}(g) \qquad (\mathrm{T} = 575\ ^\circ\mathrm{C})$

(d) $\mathrm{N_2O_3}(g) \longrightarrow \mathrm{NO}(g) + \mathrm{NO_2}(g) \qquad (\mathrm{T} = -10.0\ ^\circ\mathrm{C})$

(e) $\mathrm{SnCl_4}(l) \longrightarrow \mathrm{SnCl_4}(g) \qquad (\mathrm{T} = 200\ ^\circ\mathrm{C})$

Solution:

In each of the following, the value of Δ*G* is not given at the temperature of the reaction. Therefore, we must calculate Δ*G* from the values Δ*H*° and Δ*S* and then calculate Δ*G* from the relation Δ*G* = Δ*H*° − *T*Δ*S*°.
(a) *K* = 1.29;
(b) *K* = 2.51 × 10^{-3};
(c) *K* = 4.83 × 10^{3};
(d) *K* = 0.219;
(e) *K* = 16.1

Exercise:

Problem: Calculate the equilibrium constant at the temperature given.

(a) $\mathrm{I_2}(s) + \mathrm{Cl_2}(g) \longrightarrow 2\mathrm{ICl}(g) \qquad (\mathrm{T} = 100\ ^\circ\mathrm{C})$

(b) $\mathrm{H_2}(g) + \mathrm{I_2}(s) \longrightarrow 2\mathrm{HI}(g) \qquad (\mathrm{T} = 0.0\ ^\circ\mathrm{C})$

(c) $\mathrm{CS_2}(g) + 3\mathrm{Cl_2}(g) \longrightarrow \mathrm{CCl_4}(g) + \mathrm{S_2Cl_2}(g) \qquad (\mathrm{T} = 125\ ^\circ\mathrm{C})$

(d) $2\mathrm{SO_2}(g) + \mathrm{O_2}(g) \longrightarrow 2\mathrm{SO_3}(g) \qquad (\mathrm{T} = 675\ ^\circ\mathrm{C})$

(e) $\mathrm{CS_2}(g) \longrightarrow \mathrm{CS_2}(l) \qquad (\mathrm{T} = 90\ ^\circ\mathrm{C})$

Exercise:

Problem: Consider the following reaction at 298 K: $\mathrm{N_2O_4}(g) \rightleftharpoons 2\mathrm{NO_2}(g) \qquad K_P = 0.142$

What is the standard free energy change at this temperature? Describe what happens to the initial system, where the reactants and products are in standard states, as it approaches equilibrium.

Solution:

The standard free energy change is $\Delta G^\circ_{298} = -RT \ln K = 4.84\ \mathrm{kJ/mol}$. When reactants and products are in their standard states (1 bar or 1 atm), $Q = 1$. As the reaction proceeds toward equilibrium, the reaction shifts left (the amount of products drops while the amount of reactants increases): $Q < 1$, and ΔG_{298} becomes less positive as it approaches zero. At equilibrium, $Q = K$, and $\Delta G = 0$.

Exercise:

Problem:

Determine the normal boiling point (in kelvin) of dichloroethane, CH_2Cl_2. Find the actual boiling point using the Internet or some other source, and calculate the percent error in the temperature. Explain the differences, if any, between the two values.

Exercise:

Problem: Under what conditions is $\mathrm{N_2O_3}(g) \longrightarrow \mathrm{NO}(g) + \mathrm{NO_2}(g)$ spontaneous?

Solution:

The reaction will be spontaneous at temperatures greater than 287 K.

Exercise:

Problem:

At room temperature, the equilibrium constant (K_w) for the self-ionization of water is 1.00×10^{-14}. Using this information, calculate the standard free energy change for the aqueous reaction of hydrogen ion with hydroxide ion to produce water. (Hint: The reaction is the reverse of the self-ionization reaction.)

Exercise:

Problem:

Hydrogen sulfide is a pollutant found in natural gas. Following its removal, it is converted to sulfur by the reaction $2\mathrm{H_2S}(g) + \mathrm{SO_2}(g) \rightleftharpoons \frac{3}{8}\mathrm{S_8}(s,\ \mathrm{rhombic}) + 2\mathrm{H_2O}(l)$. What is the equilibrium constant for this reaction? Is the reaction endothermic or exothermic?

Solution:

$K = 5.35 \times 10^{15}$
The process is exothermic.

Exercise:

Problem:

Consider the decomposition of $CaCO_3(s)$ into $CaO(s)$ and $CO_2(g)$. What is the equilibrium partial pressure of CO_2 at room temperature?

Exercise:

Problem:

In the laboratory, hydrogen chloride ($HCl(g)$) and ammonia ($NH_3(g)$) often escape from bottles of their solutions and react to form the ammonium chloride ($NH_4Cl(s)$), the white glaze often seen on glassware. Assuming that the number of moles of each gas that escapes into the room is the same, what is the maximum partial pressure of HCl and NH_3 in the laboratory at room temperature? (Hint: The partial pressures will be equal and are at their maximum value when at equilibrium.)

Solution:

1.0×10^{-8} atm. This is the maximum pressure of the gases under the stated conditions.

Exercise:

Problem:

Benzene can be prepared from acetylene. $3C_2H_2(g) \rightleftharpoons C_6H_6(g)$. Determine the equilibrium constant at 25 °C and at 850 °C. Is the reaction spontaneous at either of these temperatures? Why is all acetylene not found as benzene?

Exercise:

Problem:

Carbon dioxide decomposes into CO and O_2 at elevated temperatures. What is the equilibrium partial pressure of oxygen in a sample at 1000 °C for which the initial pressure of CO_2 was 1.15 atm?

Solution:

$x = 1.29 \times 10^{-5}\ \text{atm} = P_{O_2}$

Exercise:

Problem:

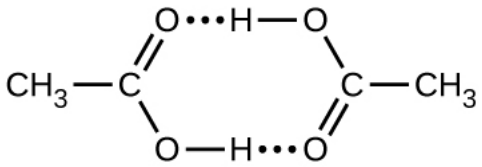
Carbon tetrachloride, an important industrial solvent, is prepared by the chlorination of methane at 850 K.
$CH_4(g) + 4Cl_2(g) \longrightarrow CCl_4(g) + 4HCl(g)$

What is the equilibrium constant for the reaction at 850 K? Would the reaction vessel need to be heated or cooled to keep the temperature of the reaction constant?

Exercise:

Problem: Acetic acid, CH_3CO_2H, can form a dimer, $(CH_3CO_2H)_2$, in the gas phase.
$2CH_3CO_2H(g) \longrightarrow (CH_3CO_2H)_2(g)$

The dimer is held together by two hydrogen bonds with a total strength of 66.5 kJ per mole of dimer.



At 25 °C, the equilibrium constant for the dimerization is 1.3×10^3 (pressure in atm). What is $\Delta S°$ for the reaction?

Solution:

−0.16 kJ

Exercise:

Problem: Nitric acid, HNO_3, can be prepared by the following sequence of reactions:
$4NH_3(g) + 5O_2(g) \longrightarrow 4NO(g) + 6H_2O(g)$
$2NO(g) + O_2(g) \longrightarrow 2NO_2(g)$
$3NO_2(g) + H_2O(l) \longrightarrow 2HNO_3(l) + NO(g)$

How much heat is evolved when 1 mol of $NH_3(g)$ is converted to $HNO_3(l)$? Assume standard states at 25 °C.

Exercise:

Problem: Determine $\Delta G°$ for the following reactions.

(a) Antimony pentachloride decomposes at 448 °C. The reaction is:
$SbCl_5(g) \longrightarrow SbCl_3(g) + Cl_2(g)$

An equilibrium mixture in a 5.00 L flask at 448 °C contains 3.85 g of $SbCl_5$, 9.14 g of $SbCl_3$, and 2.84 g of Cl_2.

(b) Chlorine molecules dissociate according to this reaction:
$Cl_2(g) \longrightarrow 2Cl(g)$

1.00% of Cl_2 molecules dissociate at 975 K and a pressure of 1.00 atm.

Solution:

(a) 22.1 kJ;
(b) 61.6 kJ/mol

Exercise:

Problem:

Given that the ΔG_f° for $Pb^{2+}(aq)$ and $Cl^-(aq)$ is −24.3 kJ/mole and −131.2 kJ/mole respectively, determine the solubility product, K_{sp}, for $PbCl_2(s)$.

Exercise:

Problem:

Determine the standard free energy change, ΔG_f°, for the formation of $S^{2-}(aq)$ given that the ΔG_f° for $Ag^+(aq)$ and $Ag_2S(s)$ are 77.1 kJ/mole and −39.5 kJ/mole respectively, and the solubility product for $Ag_2S(s)$ is 8×10^{-51}.

Solution:

90 kJ/mol

Exercise:

Problem:

Determine the standard enthalpy change, entropy change, and free energy change for the conversion of diamond to graphite. Discuss the spontaneity of the conversion with respect to the enthalpy and entropy changes. Explain why diamond spontaneously changing into graphite is not observed.

Exercise:

The evaporation of one mole of water at 298 K has a standard free energy change of 8.58 kJ.

Problem: $\mathrm{H_2O}(l) \rightleftharpoons \mathrm{H_2O}(g) \qquad \Delta G^\circ_{298} = 8.58 \text{ kJ}$

(a) Is the evaporation of water under standard thermodynamic conditions spontaneous?

(b) Determine the equilibrium constant, K_P, for this physical process.

(c) By calculating ΔG, determine if the evaporation of water at 298 K is spontaneous when the partial pressure of water, $P_{\mathrm{H_2O}}$, is 0.011 atm.

(d) If the evaporation of water were always nonspontaneous at room temperature, wet laundry would never dry when placed outside. In order for laundry to dry, what must be the value of $P_{\mathrm{H_2O}}$ in the air?

Solution:

(a) Under standard thermodynamic conditions, the evaporation is nonspontaneous; (b) K_p = 0.031; (c) The evaporation of water is spontaneous; (d) $P_{\mathrm{H_2O}}$ must always be less than K_p or less than 0.031 atm. 0.031 atm represents air saturated with water vapor at 25 °C, or 100% humidity.

Exercise:

Problem:

In glycolysis, the reaction of glucose (Glu) to form glucose-6-phosphate (G6P) requires ATP to be present as described by the following equation:

$\mathrm{Glu} + \mathrm{ATP} \longrightarrow \mathrm{G6P} + \mathrm{ADP} \qquad \Delta G^\circ_{298} = -17 \text{ kJ}$

In this process, ATP becomes ADP summarized by the following equation:

$\mathrm{ATP} \longrightarrow \mathrm{ADP} \qquad \Delta G^\circ_{298} = -30 \text{ kJ}$

Determine the standard free energy change for the following reaction, and explain why ATP is necessary to drive this process:

$\mathrm{Glu} \longrightarrow \mathrm{G6P} \qquad \Delta G^\circ_{298} = ?$

Exercise:

Problem:

One of the important reactions in the biochemical pathway glycolysis is the reaction of glucose-6-phosphate (G6P) to form fructose-6-phosphate (F6P):

$\mathrm{G6P} \rightleftharpoons \mathrm{F6P} \qquad \Delta G^\circ_{298} = 1.7 \text{ kJ}$

(a) Is the reaction spontaneous or nonspontaneous under standard thermodynamic conditions?

(b) Standard thermodynamic conditions imply the concentrations of G6P and F6P to be 1 *M*, however, in a typical cell, they are not even close to these values. Calculate ΔG when the concentrations of G6P and F6P

are 120 μ*M* and 28 μ*M* respectively, and discuss the spontaneity of the forward reaction under these conditions. Assume the temperature is 37 °C.

Solution:

(a) Nonspontaneous as $\Delta G^{\circ}_{298} > 0$; (b) $\Delta G = \Delta G^{\circ} + RT \ln Q$,
$\Delta G = 1.7 \times 10^3 + \left(8.314 \times 310 \times \ln \frac{28}{120}\right) = -2.1$ kJ. The forward reaction to produce F6P is spontaneous under these conditions.

Exercise:

Problem:

Without doing a numerical calculation, determine which of the following will reduce the free energy change for the reaction, that is, make it less positive or more negative, when the temperature is increased. Explain.

(a) $\mathrm{N_2}(g) + 3\mathrm{H_2}(g) \longrightarrow 2\mathrm{NH_3}(g)$

(b) $\mathrm{HCl}(g) + \mathrm{NH_3}(g) \longrightarrow \mathrm{NH_4Cl}(s)$

(c) $(\mathrm{NH_4})_2\mathrm{Cr_2O_7}(s) \longrightarrow \mathrm{Cr_2O_3}(s) + 4\mathrm{H_2O}(g) + \mathrm{N_2}(g)$

(d) $2\mathrm{Fe}(s) + 3\mathrm{O_2}(g) \longrightarrow \mathrm{Fe_2O_3}(s)$

Exercise:

Problem:

When ammonium chloride is added to water and stirred, it dissolves spontaneously and the resulting solution feels cold. Without doing any calculations, deduce the signs of Δ*G*, Δ*H*, and Δ*S* for this process, and justify your choices.

Solution:

Δ*G* is negative as the process is spontaneous. Δ*H* is positive as with the solution becoming cold, the dissolving must be endothermic. Δ*S* must be positive as this drives the process, and it is expected for the dissolution of any soluble ionic compound.

Exercise:

Problem:

An important source of copper is from the copper ore, chalcocite, a form of copper(I) sulfide. When heated, the Cu_2S decomposes to form copper and sulfur described by the following equation:
$\mathrm{Cu_2S}(s) \longrightarrow \mathrm{Cu}(s) + \mathrm{S}(s)$

(a) Determine ΔG°_{298} for the decomposition of $Cu_2S(s)$.

(b) The reaction of sulfur with oxygen yields sulfur dioxide as the only product. Write an equation that describes this reaction, and determine ΔG°_{298} for the process.

(c) The production of copper from chalcocite is performed by roasting the Cu_2S in air to produce the Cu. By combining the equations from Parts (a) and (b), write the equation that describes the roasting of the chalcocite, and explain why coupling these reactions together makes for a more efficient process for the production of the copper.

Exercise:

Problem:

What happens to ΔG (becomes more negative or more positive) for the following chemical reactions when the partial pressure of oxygen is increased?

(a) $\mathrm{S}(s) + \mathrm{O}_2(g) \longrightarrow \mathrm{SO}_2(g)$

(b) $2\mathrm{SO}_2(g) + \mathrm{O}_2(g) \longrightarrow 2\mathrm{SO}_3(g)$

(c) $\mathrm{HgO}(s) \longrightarrow \mathrm{Hg}(l) + \mathrm{O}_2(g)$

Solution:

(a) Increasing the oxygen partial pressure will yield a decrease in *Q* and ΔG thus becomes more negative.

(b) Increasing the oxygen partial pressure will yield a decrease in *Q* and ΔG thus becomes more negative.

(c) Increasing the oxygen partial pressure will yield an increase in *Q* and ΔG thus becomes more positive.

Glossary

Gibbs free energy change (*G*)
: thermodynamic property defined in terms of system enthalpy and entropy; all spontaneous processes involve a decrease in *G*

standard free energy change (Δ*G*°)
: change in free energy for a process occurring under standard conditions (1 bar pressure for gases, 1 M concentration for solutions)

standard free energy of formation $(\Delta G_{\mathrm{f}}^{\circ})$
: change in free energy accompanying the formation of one mole of substance from its elements in their standard states

Potential, Kinetic, Free, and Activation Energy
By the end of this section, you will be able to:

- Define “energy”
- Explain the difference between kinetic and potential energy
- Discuss the concepts of free energy and activation energy
- Describe endergonic and exergonic reactions

Energy is defined as the ability to do work. As you’ve learned, energy exists in different forms. For example, electrical energy, light energy, and heat energy are all different types of energy. While these are all familiar types of energy that one can see or feel, there is another type of energy that is much less tangible. This energy is associated with something as simple as an object held above the ground. In order to appreciate the way energy flows into and out of biological systems, it is important to understand more about the different types of energy that exist in the physical world.

Types of Energy

When an object is in motion, there is energy associated with that object. In the example of an airplane in flight, there is a great deal of energy associated with the motion of the airplane. This is because moving objects are capable of enacting a change, or doing work. Think of a wrecking ball. Even a slow-moving wrecking ball can do a great deal of damage to other objects. However, a wrecking ball that is not in motion is incapable of performing work. Energy associated with objects in motion is called **kinetic energy**. A speeding bullet, a walking person, the rapid movement of molecules in the air (which produces heat), and electromagnetic radiation like light all have kinetic energy.

Now what if that same motionless wrecking ball is lifted two stories above a car with a crane? If the suspended wrecking ball is unmoving, is there energy associated with it? The answer is yes. The suspended wrecking ball has energy associated with it that is fundamentally different from the kinetic energy of objects in motion. This form of energy results from the fact that there is the *potential* for the wrecking ball to do work. If it is released, indeed it would do work. Because this type of energy refers to the potential

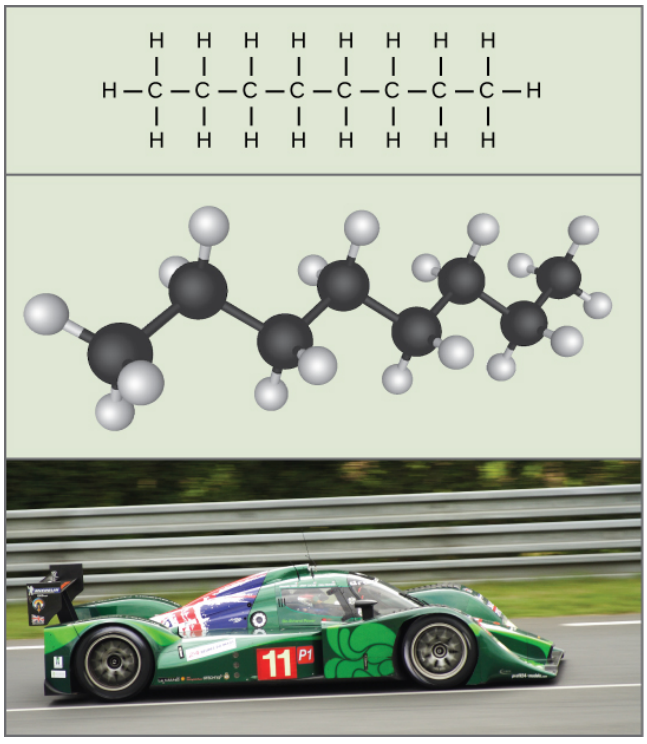
to do work, it is called **potential energy**. Objects transfer their energy between kinetic and potential in the following way: As the wrecking ball hangs motionless, it has 0 kinetic and 100 percent potential energy. Once it is released, its kinetic energy begins to increase because it builds speed due to gravity. At the same time, as it nears the ground, it loses potential energy. Somewhere mid-fall it has 50 percent kinetic and 50 percent potential energy. Just before it hits the ground, the ball has nearly lost its potential energy and has near-maximal kinetic energy. Other examples of potential energy include the energy of water held behind a dam ([link]), or a person about to skydive out of an airplane.



Water behind a dam has potential energy. Moving water, such as in a waterfall or a rapidly flowing river, has kinetic energy. (credit “dam”: modification of work by "Pascal"/Flickr; credit “waterfall”: modification of work by Frank Gualtieri)

Potential energy is not only associated with the location of matter (such as a child sitting on a tree branch), but also with the structure of matter. A spring on the ground has potential energy if it is compressed; so does a rubber band that is pulled taut. The very existence of living cells relies heavily on structural potential energy. On a chemical level, the bonds that hold the atoms of molecules together have potential energy. Remember that anabolic

cellular pathways require energy to synthesize complex molecules from simpler ones, and catabolic pathways release energy when complex molecules are broken down. The fact that energy can be released by the breakdown of certain chemical bonds implies that those bonds have potential energy. In fact, there is potential energy stored within the bonds of all the food molecules we eat, which is eventually harnessed for use. This is because these bonds can release energy when broken. The type of potential energy that exists within chemical bonds, and is released when those bonds are broken, is called **chemical energy** ([link]). Chemical energy is responsible for providing living cells with energy from food. The release of energy is brought about by breaking the molecular bonds within fuel molecules.



The molecules in gasoline (octane, the chemical formula shown) contain chemical energy within the chemical bonds. This energy is transformed into kinetic energy

that allows a car to race on a racetrack. (credit “car”: modification of work by Russell Trow)

Note:
Link to Learning

Visit this site and select “A simple pendulum” on the menu (under “Harmonic Motion”) to see the shifting kinetic (K) and potential energy (U) of a pendulum in motion.

Free Energy

After learning that chemical reactions release energy when energy-storing bonds are broken, an important next question is how is the energy associated with chemical reactions quantified and expressed? How can the energy released from one reaction be compared to that of another reaction? A measurement of **free energy** is used to quantitate these energy transfers. Free energy is called Gibbs free energy (abbreviated with the letter G) after Josiah Willard Gibbs, the scientist who developed the measurement. Recall that according to the second law of thermodynamics, all energy transfers involve the loss of some amount of energy in an unusable form such as heat, resulting in entropy. Gibbs free energy specifically refers to the energy associated with a chemical reaction that is available after entropy is

accounted for. In other words, Gibbs free energy is usable energy, or energy that is available to do work.

Every chemical reaction involves a change in free energy, called delta G (ΔG). The change in free energy can be calculated for any system that undergoes such a change, such as a chemical reaction. To calculate ΔG, subtract the amount of energy lost to entropy (denoted as ΔS) from the total energy change of the system. This total energy change in the system is called **enthalpy** and is denoted as ΔH . The formula for calculating ΔG is as follows, where the symbol T refers to absolute temperature in Kelvin (degrees Celsius + 273):

Equation:

$$\Delta \mathrm{G} = \Delta \mathrm{H} - \mathrm{T}\Delta \mathrm{S}$$

The standard free energy change of a chemical reaction is expressed as an amount of energy per mole of the reaction product (either in kilojoules or kilocalories, kJ/mol or kcal/mol; 1 kJ = 0.239 kcal) under standard pH, temperature, and pressure conditions. Standard pH, temperature, and pressure conditions are generally calculated at pH 7.0 in biological systems, 25 degrees Celsius, and 100 kilopascals (1 atm pressure), respectively. It is important to note that cellular conditions vary considerably from these standard conditions, and so standard calculated ΔG values for biological reactions will be different inside the cell.

Endergonic Reactions and Exergonic Reactions

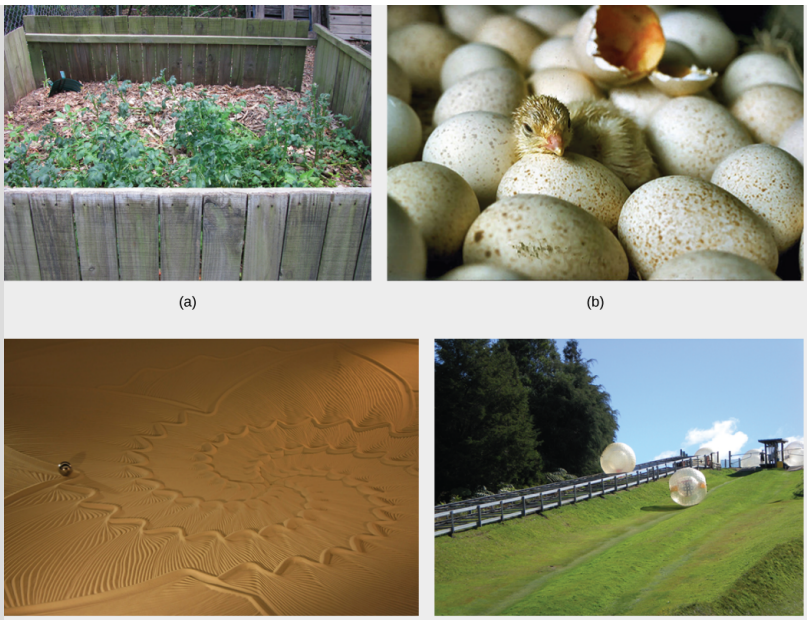
If energy is released during a chemical reaction, then the resulting value from the above equation will be a negative number. In other words, reactions that release energy have a $\Delta G < 0$. A negative ΔG also means that the products of the reaction have less free energy than the reactants, because they gave off some free energy during the reaction. Reactions that have a negative ΔG and consequently release free energy are called **exergonic reactions**. Think: *ex*ergonic means energy is *ex*iting the system. These reactions are also referred to as spontaneous reactions, because they can

occur without the addition of energy into the system. Understanding which chemical reactions are spontaneous and release free energy is extremely useful for biologists, because these reactions can be harnessed to perform work inside the cell. An important distinction must be drawn between the term spontaneous and the idea of a chemical reaction that occurs immediately. Contrary to the everyday use of the term, a spontaneous reaction is not one that suddenly or quickly occurs. The rusting of iron is an example of a spontaneous reaction that occurs slowly, little by little, over time.

If a chemical reaction requires an input of energy rather than releasing energy, then the ΔG for that reaction will be a positive value. In this case, the products have more free energy than the reactants. Thus, the products of these reactions can be thought of as energy-storing molecules. These chemical reactions are called **endergonic reactions**, and they are non-spontaneous. An endergonic reaction will not take place on its own without the addition of free energy.

Let's revisit the example of the synthesis and breakdown of the food molecule, glucose. Remember that the building of complex molecules, such as sugars, from simpler ones is an anabolic process and requires energy. Therefore, the chemical reactions involved in anabolic processes are endergonic reactions. On the other hand, the catabolic process of breaking sugar down into simpler molecules releases energy in a series of exergonic reactions. Like the example of rust above, the breakdown of sugar involves spontaneous reactions, but these reactions don't occur instantaneously. [link] shows some other examples of endergonic and exergonic reactions. Later sections will provide more information about what else is required to make even spontaneous reactions happen more efficiently.

Note:
Art Connection

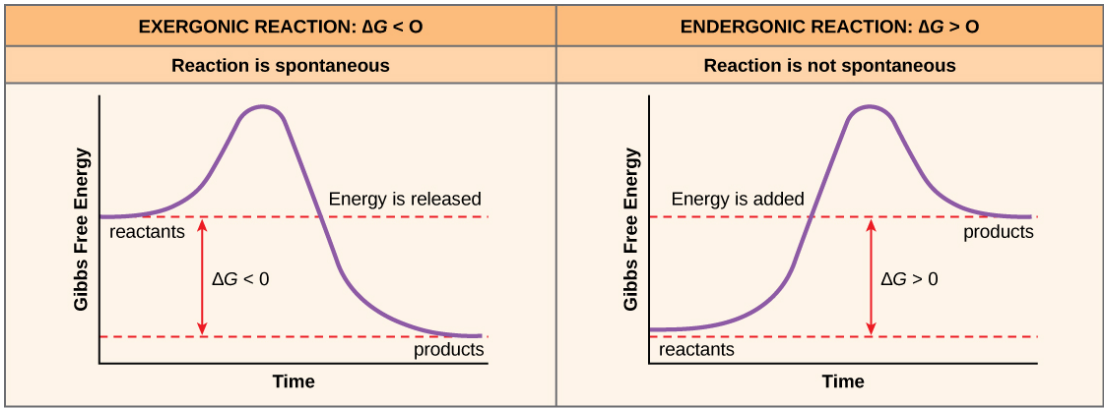


Shown are some examples of endergonic processes (ones that require energy) and exergonic processes (ones that release energy). These include (a) a compost pile decomposing, (b) a chick hatching from a fertilized egg, (c) sand art being destroyed, and (d) a ball rolling down a hill. (credit a: modification of work by Natalie Maynor; credit b: modification of work by USDA; credit c: modification of work by "Athlex"/Flickr; credit d: modification of work by Harry Malsch)

Look at each of the processes shown, and decide if it is endergonic or exergonic. In each case, does enthalpy increase or decrease, and does entropy increase or decrease?

An important concept in the study of metabolism and energy is that of chemical equilibrium. Most chemical reactions are reversible. They can

proceed in both directions, releasing energy into their environment in one direction, and absorbing it from the environment in the other direction ([link]). The same is true for the chemical reactions involved in cell metabolism, such as the breaking down and building up of proteins into and from individual amino acids, respectively. Reactants within a closed system will undergo chemical reactions in both directions until a state of equilibrium is reached. This state of equilibrium is one of the lowest possible free energy and a state of maximal entropy. Energy must be put into the system to push the reactants and products away from a state of equilibrium. Either reactants or products must be added, removed, or changed. If a cell were a closed system, its chemical reactions would reach equilibrium, and it would die because there would be insufficient free energy left to perform the work needed to maintain life. In a living cell, chemical reactions are constantly moving towards equilibrium, but never reach it. This is because a living cell is an open system. Materials pass in and out, the cell recycles the products of certain chemical reactions into other reactions, and chemical equilibrium is never reached. In this way, living organisms are in a constant energy-requiring, uphill battle against equilibrium and entropy. This constant supply of energy ultimately comes from sunlight, which is used to produce nutrients in the process of photosynthesis.



Exergonic and endergonic reactions result in changes in Gibbs free energy. Exergonic reactions release energy; endergonic reactions require energy to proceed.

Activation Energy

There is another important concept that must be considered regarding endergonic and exergonic reactions. Even exergonic reactions require a small amount of energy input to get going before they can proceed with their energy-releasing steps. These reactions have a net release of energy, but still require some energy in the beginning. This small amount of energy input necessary for all chemical reactions to occur is called the **activation energy** (or free energy of activation) and is abbreviated E_A ([link]).

Why would an energy-releasing, negative ΔG reaction actually require some energy to proceed? The reason lies in the steps that take place during a chemical reaction. During chemical reactions, certain chemical bonds are broken and new ones are formed. For example, when a glucose molecule is broken down, bonds between the carbon atoms of the molecule are broken. Since these are energy-storing bonds, they release energy when broken. However, to get them into a state that allows the bonds to break, the molecule must be somewhat contorted. A small energy input is required to achieve this contorted state. This contorted state is called the **transition state**, and it is a high-energy, unstable state. For this reason, reactant molecules don't last long in their transition state, but very quickly proceed to the next steps of the chemical reaction. Free energy diagrams illustrate the energy profiles for a given reaction. Whether the reaction is exergonic or endergonic determines whether the products in the diagram will exist at a lower or higher energy state than both the reactants and the products. However, regardless of this measure, the transition state of the reaction exists at a higher energy state than the reactants, and thus, E_A is always positive.

Note:
Link to Learning



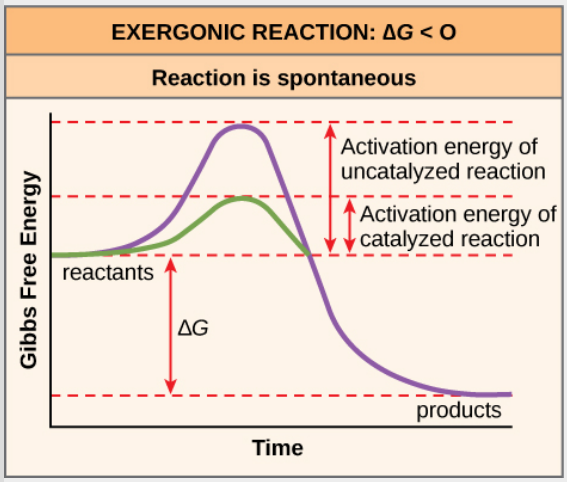
Watch an animation of the move from free energy to transition state at this site.

Where does the activation energy required by chemical reactants come from? The source of the activation energy needed to push reactions forward is typically heat energy from the surroundings. **Heat energy** (the total bond energy of reactants or products in a chemical reaction) speeds up the motion of molecules, increasing the frequency and force with which they collide; it also moves atoms and bonds within the molecule slightly, helping them reach their transition state. For this reason, heating up a system will cause chemical reactants within that system to react more frequently. Increasing the pressure on a system has the same effect. Once reactants have absorbed enough heat energy from their surroundings to reach the transition state, the reaction will proceed.

The activation energy of a particular reaction determines the rate at which it will proceed. The higher the activation energy, the slower the chemical reaction will be. The example of iron rusting illustrates an inherently slow reaction. This reaction occurs slowly over time because of its high E_A. Additionally, the burning of many fuels, which is strongly exergonic, will take place at a negligible rate unless their activation energy is overcome by sufficient heat from a spark. Once they begin to burn, however, the chemical reactions release enough heat to continue the burning process, supplying the activation energy for surrounding fuel molecules. Like these reactions outside of cells, the activation energy for most cellular reactions is too high for heat energy to overcome at efficient rates. In other words, in order for important cellular reactions to occur at appreciable rates (number of reactions per unit time), their activation energies must be lowered ([link]); this is referred to as catalysis. This is a very good thing as far as

living cells are concerned. Important macromolecules, such as proteins, DNA, and RNA, store considerable energy, and their breakdown is exergonic. If cellular temperatures alone provided enough heat energy for these exergonic reactions to overcome their activation barriers, the essential components of a cell would disintegrate.

Note:
Art Connection



Activation energy is the energy required for a reaction to proceed, and it is lower if the reaction is catalyzed. The horizontal axis of this diagram describes the sequence of events in time.

If no activation energy were required to break down sucrose (table sugar), would you be able to store it in a sugar bowl?

Section Summary

Energy comes in many different forms. Objects in motion do physical work, and kinetic energy is the energy of objects in motion. Objects that are not in motion may have the potential to do work, and thus, have potential energy. Molecules also have potential energy because the breaking of molecular bonds has the potential to release energy. Living cells depend on the harvesting of potential energy from molecular bonds to perform work. Free energy is a measure of energy that is available to do work. The free energy of a system changes during energy transfers such as chemical reactions, and this change is referred to as ΔG.

The ΔG of a reaction can be negative or positive, meaning that the reaction releases energy or consumes energy, respectively. A reaction with a negative ΔG that gives off energy is called an exergonic reaction. One with a positive ΔG that requires energy input is called an endergonic reaction. Exergonic reactions are said to be spontaneous, because their products have less energy than their reactants. The products of endergonic reactions have a higher energy state than the reactants, and so these are nonspontaneous reactions. However, all reactions (including spontaneous -ΔG reactions) require an initial input of energy in order to reach the transition state, at which they'll proceed. This initial input of energy is called the activation energy.

Art Connections

Exercise:

Problem:

[link] Look at each of the processes shown, and decide if it is endergonic or exergonic. In each case, does enthalpy increase or decrease, and does entropy increase or decrease?

Solution:

[link] A compost pile decomposing is an exergonic process; enthalpy increases (energy is released) and entropy increases (large molecules are broken down into smaller ones). A baby developing from a fertilized egg is an endergonic process; enthalpy decreases (energy is

absorbed) and entropy decreases. Sand art being destroyed is an exergonic process; there is no change in enthalpy, but entropy increases. A ball rolling downhill is an exergonic process; enthalpy decreases (energy is released), but there is no change in entropy.

Exercise:

Problem:

[link] If no activation energy were required to break down sucrose (table sugar), would you be able to store it in a sugar bowl?

Solution:

[link] No. We can store chemical energy because of the need to overcome the barrier to its breakdown.

Review Questions

Exercise:

Problem:

Consider a pendulum swinging. Which type(s) of energy is/are associated with the pendulum in the following instances: i. the moment at which it completes one cycle, just before it begins to fall back towards the other end, ii. the moment that it is in the middle between the two ends, iii. just before it reaches the end of one cycle (just before instant i.).

a. i. potential and kinetic, ii. potential and kinetic, iii. kinetic
b. i. potential, ii. potential and kinetic, iii. potential and kinetic
c. i. potential, ii. kinetic, iii. potential and kinetic
d. i. potential and kinetic, ii. kinetic iii. kinetic

Solution:

C

Exercise:

Problem:

Which of the following comparisons or contrasts between endergonic and exergonic reactions is false?

a. Endergonic reactions have a positive ΔG and exergonic reactions have a negative ΔG
b. Endergonic reactions consume energy and exergonic reactions release energy
c. Both endergonic and exergonic reactions require a small amount of energy to overcome an activation barrier
d. Endergonic reactions take place slowly and exergonic reactions take place quickly

---

Solution:

D

Exercise:

Problem:

Which of the following is the best way to judge the relative activation energies between two given chemical reactions?

a. Compare the ΔG values between the two reactions
b. Compare their reaction rates
c. Compare their ideal environmental conditions
d. Compare the spontaneity between the two reactions

Solution:

B

Free Response

Exercise:

Problem:

Explain in your own words the difference between a spontaneous reaction and one that occurs instantaneously, and what causes this difference.

Solution:

A spontaneous reaction is one that has a negative ΔG and thus releases energy. However, a spontaneous reaction need not occur quickly or suddenly like an instantaneous reaction. It may occur over long periods due to a large energy of activation, which prevents the reaction from occurring quickly.

Exercise:

Problem:

Describe the position of the transition state on a vertical energy scale, from low to high, relative to the position of the reactants and products, for both endergonic and exergonic reactions.

Solution:

The transition state is always higher in energy than the reactants and the products of a reaction (therefore, above), regardless of whether the reaction is endergonic or exergonic.

Glossary

activation energy
: energy necessary for reactions to occur

chemical energy
: potential energy in chemical bonds that is released when those bonds are broken

endergonic
: describes chemical reactions that require energy input

enthalpy
: total energy of a system

exergonic
: describes chemical reactions that release free energy

free energy
: Gibbs free energy is the usable energy, or energy that is available to do work.

heat energy
: total bond energy of reactants or products in a chemical reaction

kinetic energy
: type of energy associated with objects or particles in motion

potential energy
: type of energy that has the potential to do work; stored energy

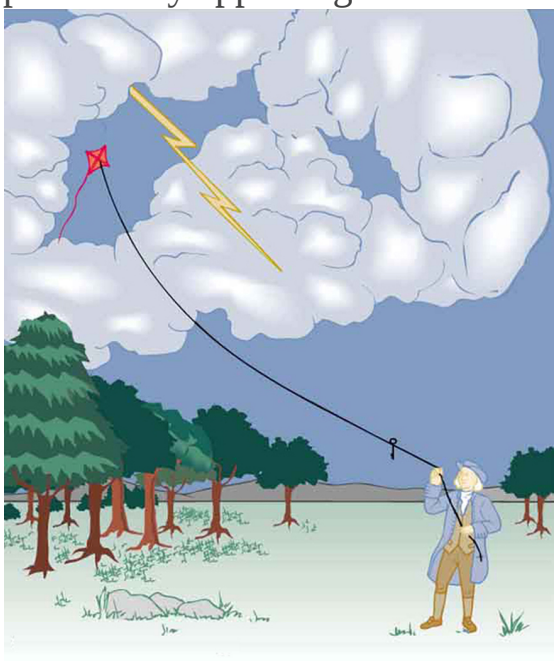
transition state
: high-energy, unstable state (an intermediate form between the substrate and the product) occurring during a chemical reaction

Introduction to Electric Charge and Electric Field
class="introduction"

Static electricity from this plastic slide causes the child’s hair to stand on end. The sliding motion stripped electrons away from the child’s body, leaving an excess of positive charges, which repel each other along each strand of hair. (credit: Ken Bosma/Wikimedia Commons)



The image of American politician and scientist Benjamin Franklin (1706–1790) flying a kite in a thunderstorm is familiar to every schoolchild. (See [link].) In this experiment, Franklin demonstrated a connection between lightning and **static electricity**. Sparks were drawn from a key hung on a kite string during an electrical storm. These sparks were like those produced by static electricity, such as the spark that jumps from your finger to a metal doorknob after you walk across a wool carpet. What Franklin demonstrated in his dangerous experiment was a connection between phenomena on two different scales: one the grand power of an electrical storm, the other an effect of more human proportions. Connections like this one reveal the underlying unity of the laws of nature, an aspect we humans find particularly appealing.



When Benjamin Franklin demonstrated that lightning was related to static electricity, he made a connection that is now part of the evidence that all directly experienced forces except the gravitational force are manifestations of the electromagnetic force.

Much has been written about Franklin. His experiments were only part of the life of a man who was a scientist, inventor, revolutionary, statesman, and writer. Franklin's experiments were not performed in isolation, nor were they the only ones to reveal connections.

For example, the Italian scientist Luigi Galvani (1737–1798) performed a series of experiments in which static electricity was used to stimulate contractions of leg muscles of dead frogs, an effect already known in humans subjected to static discharges. But Galvani also found that if he joined two metal wires (say copper and zinc) end to end and touched the other ends to muscles, he produced the same effect in frogs as static discharge. Alessandro Volta (1745–1827), partly inspired by Galvani's work, experimented with various combinations of metals and developed the battery.

During the same era, other scientists made progress in discovering fundamental connections. The periodic table was developed as the systematic properties of the elements were discovered. This influenced the development and refinement of the concept of atoms as the basis of matter. Such submicroscopic descriptions of matter also help explain a great deal more.

Atomic and molecular interactions, such as the forces of friction, cohesion, and adhesion, are now known to be manifestations of the **electromagnetic force**. Static electricity is just one aspect of the electromagnetic force, which also includes moving electricity and magnetism.

All the macroscopic forces that we experience directly, such as the sensations of touch and the tension in a rope, are due to the electromagnetic force, one of the four fundamental forces in nature. The gravitational force, another fundamental force, is actually sensed through the electromagnetic interaction of molecules, such as between those in our feet and those on the top of a bathroom scale. (The other two fundamental forces, the strong nuclear force and the weak nuclear force, cannot be sensed on the human scale.)

This chapter begins the study of electromagnetic phenomena at a fundamental level. The next several chapters will cover static electricity, moving electricity, and magnetism—collectively known as electromagnetism. In this chapter, we begin with the study of electric phenomena due to charges that are at least temporarily stationary, called electrostatics, or static electricity.

Glossary

static electricity
: a buildup of electric charge on the surface of an object

electromagnetic force
: one of the four fundamental forces of nature; the electromagnetic force consists of static electricity, moving electricity and magnetism

Static Electricity and Charge: Conservation of Charge

- Define electric charge, and describe how the two types of charge interact.
- Describe three common situations that generate static electricity.
- State the law of conservation of charge.



Borneo amber was mined in Sabah, Malaysia, from shale-sandstone-mudstone veins. When a piece of amber is rubbed with a piece of silk, the amber gains more electrons, giving it a net negative charge. At the same time, the silk, having lost electrons, becomes positively charged. (credit: Sebakoamber, Wikimedia Commons)

What makes plastic wrap cling? Static electricity. Not only are applications of static electricity common these days, its existence has been known since ancient times. The first record of its effects dates to ancient Greeks who noted more than 500 years B.C. that polishing amber temporarily enabled it

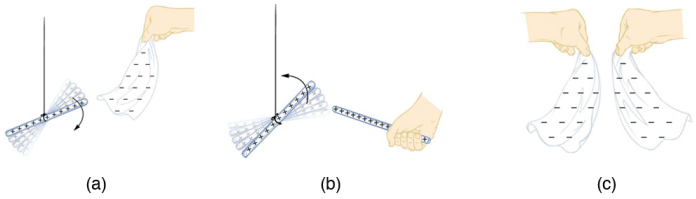
to attract bits of straw (see [link]). The very word *electric* derives from the Greek word for amber (*electron*).

Many of the characteristics of static electricity can be explored by rubbing things together. Rubbing creates the spark you get from walking across a wool carpet, for example. Static cling generated in a clothes dryer and the attraction of straw to recently polished amber also result from rubbing. Similarly, lightning results from air movements under certain weather conditions. You can also rub a balloon on your hair, and the static electricity created can then make the balloon cling to a wall. We also have to be cautious of static electricity, especially in dry climates. When we pump gasoline, we are warned to discharge ourselves (after sliding across the seat) on a metal surface before grabbing the gas nozzle. Attendants in hospital operating rooms must wear booties with aluminum foil on the bottoms to avoid creating sparks which may ignite the oxygen being used.

Some of the most basic characteristics of static electricity include:

- The effects of static electricity are explained by a physical quantity not previously introduced, called electric charge.
- There are only two types of charge, one called positive and the other called negative.
- Like charges repel, whereas unlike charges attract.
- The force between charges decreases with distance.

How do we know there are two types of **electric charge**? When various materials are rubbed together in controlled ways, certain combinations of materials always produce one type of charge on one material and the opposite type on the other. By convention, we call one type of charge “positive”, and the other type “negative.” For example, when glass is rubbed with silk, the glass becomes positively charged and the silk negatively charged. Since the glass and silk have opposite charges, they attract one another like clothes that have rubbed together in a dryer. Two glass rods rubbed with silk in this manner will repel one another, since each rod has positive charge on it. Similarly, two silk cloths so rubbed will repel, since both cloths have negative charge. [link] shows how these simple materials can be used to explore the nature of the force between charges.



A glass rod becomes positively charged when rubbed with silk, while the silk becomes negatively charged. (a) The glass rod is attracted to the silk because their charges are opposite. (b) Two similarly charged glass rods repel. (c) Two similarly charged silk cloths repel.

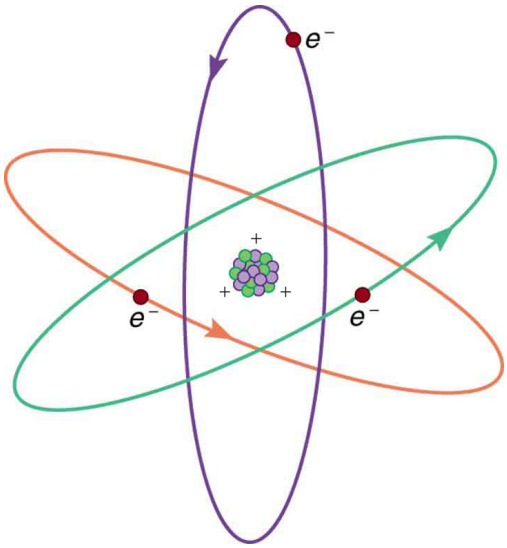
More sophisticated questions arise. Where do these charges come from? Can you create or destroy charge? Is there a smallest unit of charge? Exactly how does the force depend on the amount of charge and the distance between charges? Such questions obviously occurred to Benjamin Franklin and other early researchers, and they interest us even today.

Charge Carried by Electrons and Protons

Franklin wrote in his letters and books that he could see the effects of electric charge but did not understand what caused the phenomenon. Today we have the advantage of knowing that normal matter is made of atoms, and that atoms contain positive and negative charges, usually in equal amounts.

[link] shows a simple model of an atom with negative **electrons** orbiting its positive nucleus. The nucleus is positive due to the presence of positively charged **protons**. Nearly all charge in nature is due to electrons and protons, which are two of the three building blocks of most matter. (The third is the neutron, which is neutral, carrying no charge.) Other charge-carrying particles are observed in cosmic rays and nuclear decay, and are created in

particle accelerators. All but the electron and proton survive only a short time and are quite rare by comparison.



This simplified (and not to scale) view of an atom is called the planetary model of the atom. Negative electrons orbit a much heavier positive nucleus, as the planets orbit the much heavier sun. There the similarity ends, because forces in the atom are electromagnetic, whereas those in the planetary system are gravitational. Normal macroscopic amounts of matter contain immense numbers of atoms and molecules and, hence, even greater numbers of individual

negative and positive
charges.

The charges of electrons and protons are identical in magnitude but opposite in sign. Furthermore, all charged objects in nature are integral multiples of this basic quantity of charge, meaning that all charges are made of combinations of a basic unit of charge. Usually, charges are formed by combinations of electrons and protons. The magnitude of this basic charge is

Equation:

$$| q_e | = 1.60 \times 10^{-19} \text{ C}.$$

The symbol q is commonly used for charge and the subscript e indicates the charge of a single electron (or proton).

The SI unit of charge is the coulomb (C). The number of protons needed to make a charge of 1.00 C is

Equation:

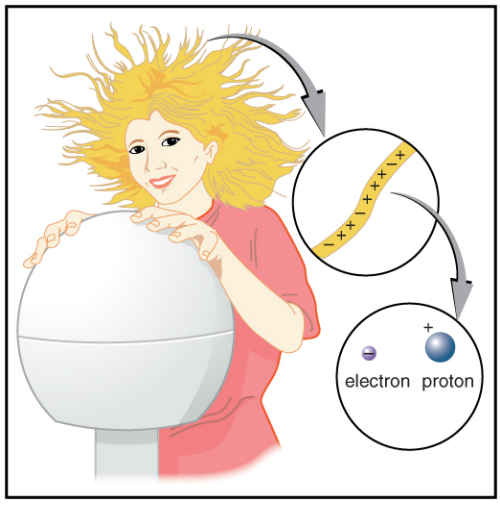
$$1.00 \text{ C} \times \frac{1 \text{ proton}}{1.60 \times 10^{-19} \text{ C}} = 6.25 \times 10^{18} \text{ protons}.$$

Similarly, 6.25×10^{18} electrons have a combined charge of −1.00 coulomb. Just as there is a smallest bit of an element (an atom), there is a smallest bit of charge. There is no directly observed charge smaller than $| q_e |$ (see Things Great and Small: The Submicroscopic Origin of Charge), and all observed charges are integral multiples of $| q_e |$.

Note:
Things Great and Small: The Submicroscopic Origin of Charge

With the exception of exotic, short-lived particles, all charge in nature is carried by electrons and protons. Electrons carry the charge we have named negative. Protons carry an equal-magnitude charge that we call positive. (See [link].) Electron and proton charges are considered fundamental building blocks, since all other charges are integral multiples of those carried by electrons and protons. Electrons and protons are also two of the three fundamental building blocks of ordinary matter. The neutron is the third and has zero total charge.

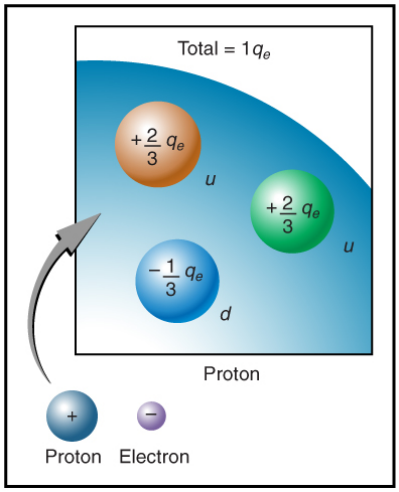
[link] shows a person touching a Van de Graaff generator and receiving excess positive charge. The expanded view of a hair shows the existence of both types of charges but an excess of positive. The repulsion of these positive like charges causes the strands of hair to repel other strands of hair and to stand up. The further blowup shows an artist's conception of an electron and a proton perhaps found in an atom in a strand of hair.



When this person touches a Van de Graaff generator, she receives an excess of positive charge, causing her hair to stand on end. The charges in

one hair are shown. An artist's conception of an electron and a proton illustrate the particles carrying the negative and positive charges. We cannot really see these particles with visible light because they are so small (the electron seems to be an infinitesimal point), but we know a great deal about their measurable properties, such as the charges they carry.

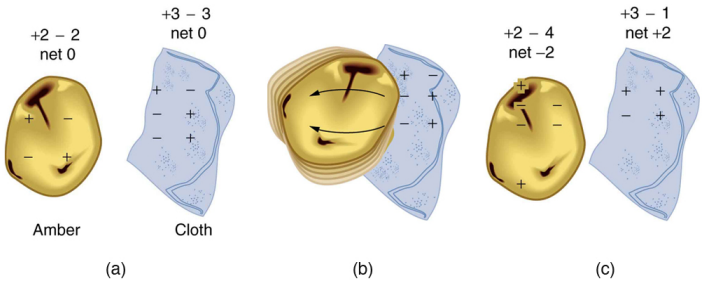
The electron seems to have no substructure; in contrast, when the substructure of protons is explored by scattering extremely energetic electrons from them, it appears that there are point-like particles inside the proton. These sub-particles, named quarks, have never been directly observed, but they are believed to carry fractional charges as seen in [link]. Charges on electrons and protons and all other directly observable particles are unitary, but these quark substructures carry charges of either $-\frac{1}{3}$ or $+\frac{2}{3}$. There are continuing attempts to observe fractional charge directly and to learn of the properties of quarks, which are perhaps the ultimate substructure of matter.



Artist’s conception of fractional quark charges inside a proton. A group of three quark charges add up to the single positive charge on the proton: $-\frac{1}{3}q_e + \frac{2}{3}q_e + \frac{2}{3}q_e = +1q_e$.

Separation of Charge in Atoms

Charges in atoms and molecules can be separated—for example, by rubbing materials together. Some atoms and molecules have a greater affinity for electrons than others and will become negatively charged by close contact in rubbing, leaving the other material positively charged. (See [link].) Positive charge can similarly be induced by rubbing. Methods other than rubbing can also separate charges. Batteries, for example, use combinations of substances that interact in such a way as to separate charges. Chemical interactions may transfer negative charge from one substance to the other, making one battery terminal negative and leaving the first one positive.



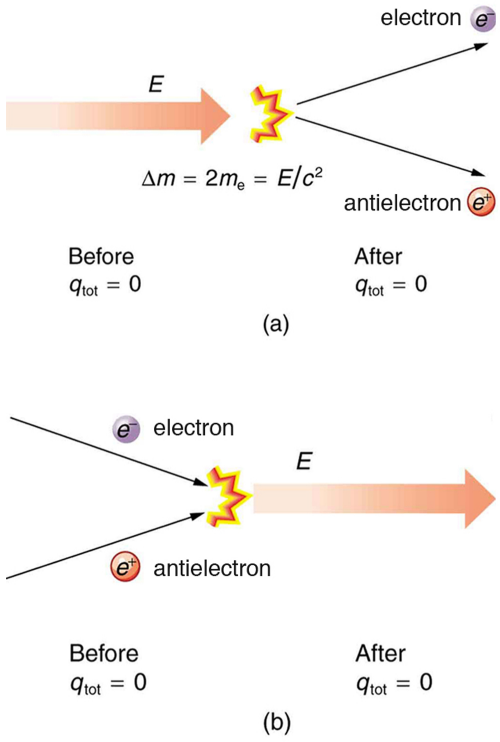
When materials are rubbed together, charges can be separated, particularly if one material has a greater affinity for electrons than another. (a) Both the amber and cloth are originally neutral, with equal positive and negative charges. Only a tiny fraction of the charges are involved, and only a few of them are shown here. (b) When rubbed together, some negative charge is transferred to the amber, leaving the cloth with a net positive charge. (c) When separated, the amber and cloth now have net charges, but the absolute value of the net positive and negative charges will be equal.

No charge is actually created or destroyed when charges are separated as we have been discussing. Rather, existing charges are moved about. In fact, in all situations the total amount of charge is always constant. This universally obeyed law of nature is called the **law of conservation of charge**.

Note:
Law of Conservation of Charge
Total charge is constant in any process.

In more exotic situations, such as in particle accelerators, mass, Δm, can be created from energy in the amount $\Delta m = \frac{E}{c^2}$. Sometimes, the created mass is charged, such as when an electron is created. Whenever a charged particle is created, another having an opposite charge is always created along with it, so that the total charge created is zero. Usually, the two particles are “matter-antimatter” counterparts. For example, an antielectron would usually be created at the same time as an electron. The antielectron has a positive charge (it is called a positron), and so the total charge created is zero. (See [link].) All particles have antimatter counterparts with opposite signs. When matter and antimatter counterparts are brought together, they completely annihilate one another. By annihilate, we mean that the mass of the two particles is converted to energy *E*, again obeying the relationship $\Delta m = \frac{E}{c^2}$. Since the two particles have equal and opposite charge, the total charge is zero before and after the annihilation; thus, total charge is conserved.

Note:
Making Connections: Conservation Laws
Only a limited number of physical quantities are universally conserved. Charge is one—energy, momentum, and angular momentum are others. Because they are conserved, these physical quantities are used to explain more phenomena and form more connections than other, less basic quantities. We find that conserved quantities give us great insight into the rules followed by nature and hints to the organization of nature. Discoveries of conservation laws have led to further discoveries, such as the weak nuclear force and the quark substructure of protons and other particles.



(a) When enough energy is present, it can be converted into matter. Here the matter created is an electron–antielectron pair. (m_e is the electron's mass.) The total charge before and after this event is zero. (b) When matter and antimatter collide, they annihilate each other; the total charge is conserved at zero before and after the annihilation.

The law of conservation of charge is absolute—it has never been observed to be violated. Charge, then, is a special physical quantity, joining a very

short list of other quantities in nature that are always conserved. Other conserved quantities include energy, momentum, and angular momentum.

Note:
PhET Explorations: Balloons and Static Electricity
Why does a balloon stick to your sweater? Rub a balloon on a sweater, then let go of the balloon and it flies over and sticks to the sweater. View the charges in the sweater, balloons, and the wall.
https://phet.colorado.edu/sims/html/balloons-and-static-electricity/latest/balloons-and-static-electricity_en.html

Section Summary

- There are only two types of charge, which we call positive and negative.
- Like charges repel, unlike charges attract, and the force between charges decreases with the square of the distance.
- The vast majority of positive charge in nature is carried by protons, while the vast majority of negative charge is carried by electrons.
- The electric charge of one electron is equal in magnitude and opposite in sign to the charge of one proton.
- An ion is an atom or molecule that has nonzero total charge due to having unequal numbers of electrons and protons.
- The SI unit for charge is the coulomb (C), with protons and electrons having charges of opposite sign but equal magnitude; the magnitude of this basic charge $|\ q_e\ |$ is
  Equation:

$$|\ q_e\ | = 1.60 \times 10^{-19}\ \text{C}.$$

- Whenever charge is created or destroyed, equal amounts of positive and negative are involved.
- Most often, existing charges are separated from neutral objects to obtain some net charge.

- Both positive and negative charges exist in neutral objects and can be separated by rubbing one object with another. For macroscopic objects, negatively charged means an excess of electrons and positively charged means a depletion of electrons.
- The law of conservation of charge ensures that whenever a charge is created, an equal charge of the opposite sign is created at the same time.

Conceptual Questions

Exercise:

Problem:

There are very large numbers of charged particles in most objects. Why, then, don’t most objects exhibit static electricity?

Exercise:

Problem:

Why do most objects tend to contain nearly equal numbers of positive and negative charges?

Problems & Exercises

Exercise:

Problem:

Common static electricity involves charges ranging from nanocoulombs to microcoulombs. (a) How many electrons are needed to form a charge of $-2.00\ \text{nC}$ (b) How many electrons must be removed from a neutral object to leave a net charge of $0.500\ \mu\text{C}$?

Solution:

(a) 1.25×10^{10}

(b) 3.13×10^{12}

Exercise:

Problem:

If 1.80×10^{20} electrons move through a pocket calculator during a full day's operation, how many coulombs of charge moved through it?

Exercise:

Problem:

To start a car engine, the car battery moves 3.75×10^{21} electrons through the starter motor. How many coulombs of charge were moved?

Solution:

-600 C

Exercise:

Problem:

A certain lightning bolt moves 40.0 C of charge. How many fundamental units of charge $\left| q_e \right|$ is this?

Glossary

electric charge
: a physical property of an object that causes it to be attracted toward or repelled from another charged object; each charged object generates and is influenced by a force called an electromagnetic force

law of conservation of charge
: states that whenever a charge is created, an equal amount of charge with the opposite sign is created simultaneously

electron

a particle orbiting the nucleus of an atom and carrying the smallest unit of negative charge

proton
: a particle in the nucleus of an atom and carrying a positive charge equal in magnitude and opposite in sign to the amount of negative charge carried by an electron

Conductors and Insulators

- Define conductor and insulator, explain the difference, and give examples of each.
- Describe three methods for charging an object.
- Explain what happens to an electric force as you move farther from the source.
- Define polarization.

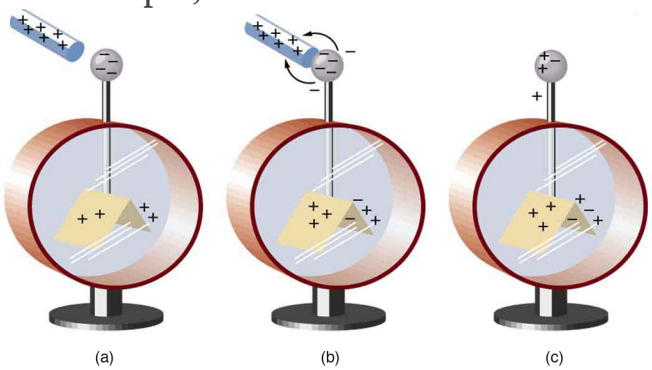


This power adapter uses metal wires and connectors to conduct electricity from the wall socket to a laptop computer. The conducting wires allow electrons to move freely through the cables, which are shielded by rubber and plastic. These materials act as insulators that don’t allow electric charge to escape outward. (credit: Evan-Amos, Wikimedia Commons)

Some substances, such as metals and salty water, allow charges to move through them with relative ease. Some of the electrons in metals and similar conductors are not bound to individual atoms or sites in the material. These **free electrons** can move through the material much as air moves through loose sand. Any substance that has free electrons and allows charge to move

relatively freely through it is called a **conductor**. The moving electrons may collide with fixed atoms and molecules, losing some energy, but they can move in a conductor. Superconductors allow the movement of charge without any loss of energy. Salty water and other similar conducting materials contain free ions that can move through them. An ion is an atom or molecule having a positive or negative (nonzero) total charge. In other words, the total number of electrons is not equal to the total number of protons.

Other substances, such as glass, do not allow charges to move through them. These are called **insulators**. Electrons and ions in insulators are bound in the structure and cannot move easily—as much as 10^{23} times more slowly than in conductors. Pure water and dry table salt are insulators, for example, whereas molten salt and salty water are conductors.



An electroscope is a favorite instrument in physics demonstrations and student laboratories. It is typically made with gold foil leaves hung from a (conducting) metal stem and is insulated from the room air in a glass-walled container. (a) A positively charged glass rod is brought near the tip of the electroscope, attracting electrons to the top and leaving a net positive charge on the leaves. Like charges in the light flexible gold leaves

repel, separating them. (b) When the rod is touched against the ball, electrons are attracted and transferred, reducing the net charge on the glass rod but leaving the electroscope positively charged. (c) The excess charges are evenly distributed in the stem and leaves of the electroscope once the glass rod is removed.

Charging by Contact

[link] shows an electroscope being charged by touching it with a positively charged glass rod. Because the glass rod is an insulator, it must actually touch the electroscope to transfer charge to or from it. (Note that the extra positive charges reside on the surface of the glass rod as a result of rubbing it with silk before starting the experiment.) Since only electrons move in metals, we see that they are attracted to the top of the electroscope. There, some are transferred to the positive rod by touch, leaving the electroscope with a net positive charge.

Electrostatic repulsion in the leaves of the charged electroscope separates them. The electrostatic force has a horizontal component that results in the leaves moving apart as well as a vertical component that is balanced by the gravitational force. Similarly, the electroscope can be negatively charged by contact with a negatively charged object.

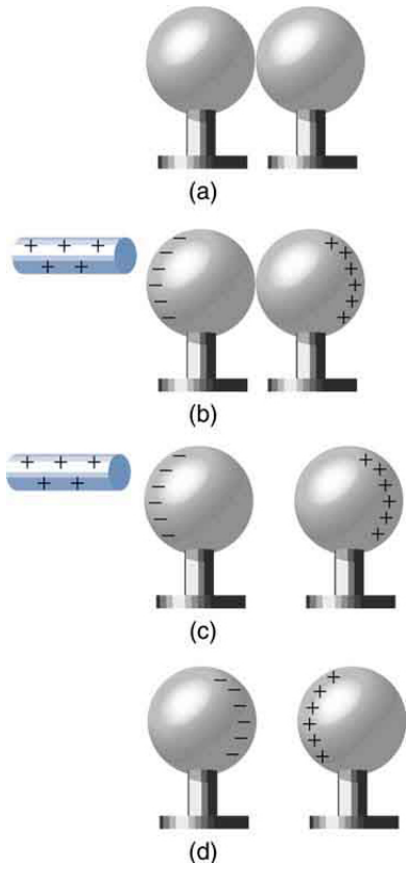
Charging by Induction

It is not necessary to transfer excess charge directly to an object in order to charge it. [link] shows a method of **induction** wherein a charge is created in a nearby object, without direct contact. Here we see two neutral metal spheres in contact with one another but insulated from the rest of the world.

A positively charged rod is brought near one of them, attracting negative charge to that side, leaving the other sphere positively charged.

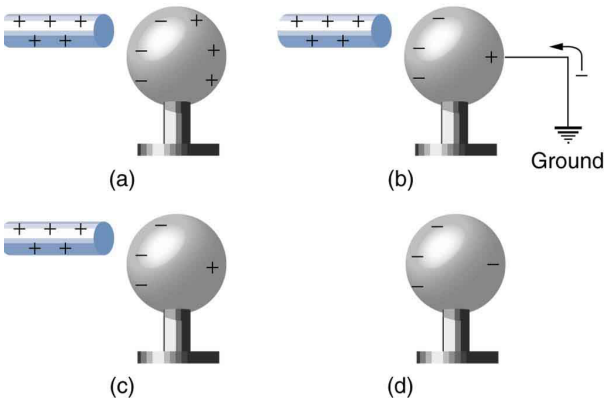
This is an example of induced **polarization** of neutral objects. Polarization is the separation of charges in an object that remains neutral. If the spheres are now separated (before the rod is pulled away), each sphere will have a net charge. Note that the object closest to the charged rod receives an opposite charge when charged by induction. Note also that no charge is removed from the charged rod, so that this process can be repeated without depleting the supply of excess charge.

Another method of charging by induction is shown in [link]. The neutral metal sphere is polarized when a charged rod is brought near it. The sphere is then grounded, meaning that a conducting wire is run from the sphere to the ground. Since the earth is large and most ground is a good conductor, it can supply or accept excess charge easily. In this case, electrons are attracted to the sphere through a wire called the ground wire, because it supplies a conducting path to the ground. The ground connection is broken before the charged rod is removed, leaving the sphere with an excess charge opposite to that of the rod. Again, an opposite charge is achieved when charging by induction and the charged rod loses none of its excess charge.



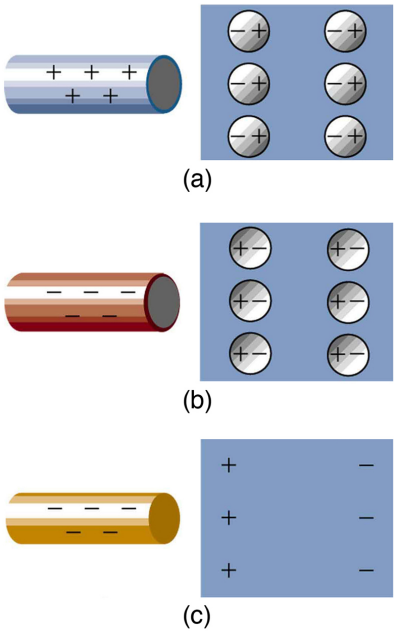
Charging by induction. (a) Two uncharged or neutral metal spheres are in contact with each other but insulated from the rest of the world. (b) A positively charged glass rod is brought near the sphere on the left, attracting negative charge and leaving the other sphere positively charged. (c) The

spheres are separated before the rod is removed, thus separating negative and positive charge. (d) The spheres retain net charges after the inducing rod is removed—without ever having been touched by a charged object.



Charging by induction, using a ground connection. (a) A positively charged rod is brought near a neutral metal sphere, polarizing it. (b) The sphere is grounded, allowing electrons to be attracted from the earth’s ample supply. (c) The ground connection is broken. (d) The positive rod is

removed, leaving the sphere with an induced negative charge.



Both positive and negative objects attract a neutral object by polarizing its molecules. (a) A positive object brought near a neutral insulator polarizes its molecules. There is a slight shift in the distribution of the electrons orbiting the molecule, with

unlike charges being brought nearer and like charges moved away. Since the electrostatic force decreases with distance, there is a net attraction. (b) A negative object produces the opposite polarization, but again attracts the neutral object. (c) The same effect occurs for a conductor; since the unlike charges are closer, there is a net attraction.

Neutral objects can be attracted to any charged object. The pieces of straw attracted to polished amber are neutral, for example. If you run a plastic comb through your hair, the charged comb can pick up neutral pieces of paper. [link] shows how the polarization of atoms and molecules in neutral objects results in their attraction to a charged object.

When a charged rod is brought near a neutral substance, an insulator in this case, the distribution of charge in atoms and molecules is shifted slightly. Opposite charge is attracted nearer the external charged rod, while like charge is repelled. Since the electrostatic force decreases with distance, the repulsion of like charges is weaker than the attraction of unlike charges, and so there is a net attraction. Thus a positively charged glass rod attracts neutral pieces of paper, as will a negatively charged rubber rod. Some

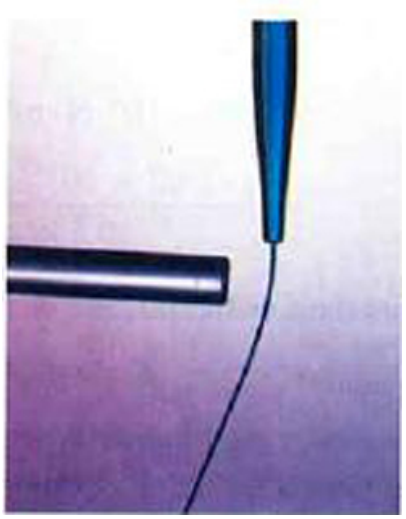
molecules, like water, are polar molecules. Polar molecules have a natural or inherent separation of charge, although they are neutral overall. Polar molecules are particularly affected by other charged objects and show greater polarization effects than molecules with naturally uniform charge distributions.

Exercise:

Check Your Understanding

Problem:

Can you explain the attraction of water to the charged rod in the figure below?



Solution:

Answer

Water molecules are polarized, giving them slightly positive and slightly negative sides. This makes water even more susceptible to a charged rod's attraction. As the water flows downward, due to the force of gravity, the charged conductor exerts a net attraction to the opposite charges in the stream of water, pulling it closer.

Note:

PhET Explorations: John Travoltage

Make sparks fly with John Travoltage. Wiggle Johnnie's foot and he picks up charges from the carpet. Bring his hand close to the door knob and get rid of the excess charge.
https://phet.colorado.edu/sims/html/john-travoltage/latest/john-travoltage_en.html

Section Summary

- Polarization is the separation of positive and negative charges in a neutral object.
- A conductor is a substance that allows charge to flow freely through its atomic structure.
- An insulator holds charge within its atomic structure.
- Objects with like charges repel each other, while those with unlike charges attract each other.
- A conducting object is said to be grounded if it is connected to the Earth through a conductor. Grounding allows transfer of charge to and from the earth’s large reservoir.
- Objects can be charged by contact with another charged object and obtain the same sign charge.
- If an object is temporarily grounded, it can be charged by induction, and obtains the opposite sign charge.
- Polarized objects have their positive and negative charges concentrated in different areas, giving them a non-symmetrical charge.
- Polar molecules have an inherent separation of charge.

Conceptual Questions

Exercise:

Problem:

An eccentric inventor attempts to levitate by first placing a large negative charge on himself and then putting a large positive charge on the ceiling of his workshop. Instead, while attempting to place a large negative charge on himself, his clothes fly off. Explain.

Exercise:

Problem:

If you have charged an electroscope by contact with a positively charged object, describe how you could use it to determine the charge of other objects. Specifically, what would the leaves of the electroscope do if other charged objects were brought near its knob?

Exercise:

Problem:

When a glass rod is rubbed with silk, it becomes positive and the silk becomes negative—yet both attract dust. Does the dust have a third type of charge that is attracted to both positive and negative? Explain.

Exercise:

Problem:

Why does a car always attract dust right after it is polished? (Note that car wax and car tires are insulators.)

Exercise:

Problem:

Describe how a positively charged object can be used to give another object a negative charge. What is the name of this process?

Exercise:

Problem:

What is grounding? What effect does it have on a charged conductor? On a charged insulator?

Problems & Exercises

Exercise:

Problem:

Suppose a speck of dust in an electrostatic precipitator has 1.0000×10^{12} protons in it and has a net charge of –5.00 nC (a very large charge for a small speck). How many electrons does it have?

Solution:

1.03×10^{12}

Exercise:

Problem:

An amoeba has 1.00×10^{16} protons and a net charge of 0.300 pC. (a) How many fewer electrons are there than protons? (b) If you paired them up, what fraction of the protons would have no electrons?

Exercise:

Problem:

A 50.0 g ball of copper has a net charge of $2.00\ \mu\text{C}$. What fraction of the copper's electrons has been removed? (Each copper atom has 29 protons, and copper has an atomic mass of 63.5.)

Solution:

9.09×10^{-13}

Exercise:

Problem:

What net charge would you place on a 100 g piece of sulfur if you put an extra electron on 1 in 10^{12} of its atoms? (Sulfur has an atomic mass of 32.1.)

Exercise:

Problem:

How many coulombs of positive charge are there in 4.00 kg of plutonium, given its atomic mass is 244 and that each plutonium atom has 94 protons?

Solution:

$1.48 \times 10^8\ \mathrm{C}$

Glossary

free electron
: an electron that is free to move away from its atomic orbit

conductor
: a material that allows electrons to move separately from their atomic orbits

insulator
: a material that holds electrons securely within their atomic orbits

grounded
: when a conductor is connected to the Earth, allowing charge to freely flow to and from Earth’s unlimited reservoir

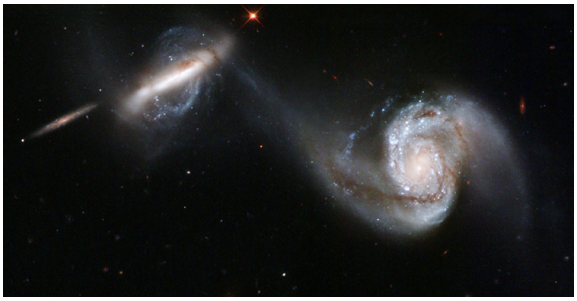
induction
: the process by which an electrically charged object brought near a neutral object creates a charge in that object

polarization
: slight shifting of positive and negative charges to opposite sides of an atom or molecule

electrostatic repulsion
: the phenomenon of two objects with like charges repelling each other

Coulomb’s Law

- State Coulomb’s law in terms of how the electrostatic force changes with the distance between two objects.
- Calculate the electrostatic force between two charged point forces, such as electrons or protons.
- Compare the electrostatic force to the gravitational attraction for a proton and an electron; for a human and the Earth.



This NASA image of Arp 87 shows the result of a strong gravitational attraction between two galaxies. In contrast, at the subatomic level, the electrostatic attraction between two objects, such as an electron and a proton, is far greater than their mutual attraction due to gravity. (credit: NASA/HST)

Through the work of scientists in the late 18th century, the main features of the **electrostatic force**—the existence of two types of charge, the observation that like charges repel, unlike charges attract, and the decrease of force with distance—were eventually refined, and expressed as a mathematical formula. The mathematical formula for the electrostatic force is called **Coulomb’s law** after the French physicist Charles Coulomb (1736–1806), who performed experiments and first proposed a formula to calculate it.

Note:
Coulomb’s Law
Equation:

$$F = k\frac{|q_1 q_2|}{r^2}.$$

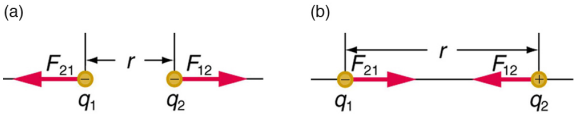
Coulomb's law calculates the magnitude of the force F between two point charges, q_1 and q_2, separated by a distance r. In SI units, the constant k is equal to

Equation:

$$k = 8.988 \times 10^9 \frac{\text{N} \cdot \text{m}^2}{\text{C}^2} \approx 8.99 \times 10^9 \frac{\text{N} \cdot \text{m}^2}{\text{C}^2}.$$

The electrostatic force is a vector quantity and is expressed in units of newtons. The force is understood to be along the line joining the two charges. (See [link].)

Although the formula for Coulomb's law is simple, it was no mean task to prove it. The experiments Coulomb did, with the primitive equipment then available, were difficult. Modern experiments have verified Coulomb's law to great precision. For example, it has been shown that the force is inversely proportional to distance between two objects squared $\left(F \propto 1/r^2\right)$ to an accuracy of 1 part in 10^{16}. No exceptions have ever been found, even at the small distances within the atom.



The magnitude of the electrostatic force F between point charges q_1 and q_2 separated by a distance r is given by Coulomb's law. Note that Newton's third law (every force exerted creates an equal and opposite force) applies as usual—the force on q_1 is equal in magnitude and opposite in direction to the force it exerts on q_2. (a) Like charges. (b) Unlike charges.

Example:

How Strong is the Coulomb Force Relative to the Gravitational Force?

Compare the electrostatic force between an electron and proton separated by 0.530×10^{-10} m with the gravitational force between them. This distance is their average separation in a hydrogen atom.

Strategy

To compare the two forces, we first compute the electrostatic force using Coulomb's law, $F = k\frac{|q_1 q_2|}{r^2}$. We then calculate the gravitational force using Newton's universal law of

gravitation. Finally, we take a ratio to see how the forces compare in magnitude.

Solution

Entering the given and known information about the charges and separation of the electron and proton into the expression of Coulomb's law yields

Equation:

$$F = k\frac{|q_1 q_2|}{r^2}$$

Equation:

$$= \left(8.99 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2\right) \times \frac{(1.60 \times 10^{-19} \text{ C})(1.60 \times 10^{-19} \text{ C})}{(0.530 \times 10^{-10} \text{ m})^2}$$

Thus the Coulomb force is

Equation:

$$F = 8.19 \times 10^{-8} \text{ N}.$$

The charges are opposite in sign, so this is an attractive force. This is a very large force for an electron—it would cause an acceleration of $8.99 \times 10^{22} \text{ m/s}^2$(verification is left as an end-of-section problem).The gravitational force is given by Newton's law of gravitation as:

Equation:

$$F_G = G\frac{mM}{r^2},$$

where $G = 6.67 \times 10^{-11} \text{ N} \cdot \text{m}^2/\text{kg}^2$. Here m and M represent the electron and proton masses, which can be found in the appendices. Entering values for the knowns yields

Equation:

$$F_G = (6.67 \times 10^{-11} \text{ N} \cdot \text{m}^2/\text{kg}^2) \times \frac{(9.11 \times 10^{-31} \text{ kg})(1.67 \times 10^{-27} \text{ kg})}{(0.530 \times 10^{-10} \text{ m})^2} = 3.61 \times 10^{-47} \text{ N}$$

This is also an attractive force, although it is traditionally shown as positive since gravitational force is always attractive. The ratio of the magnitude of the electrostatic force to gravitational force in this case is, thus,

Equation:

$$\frac{F}{F_G} = 2.27 \times 10^{39}.$$

Discussion

This is a remarkably large ratio! Note that this will be the ratio of electrostatic force to gravitational force for an electron and a proton at any distance (taking the ratio before entering numerical values shows that the distance cancels). This ratio gives some indication

of just how much larger the Coulomb force is than the gravitational force between two of the most common particles in nature.

As the example implies, gravitational force is completely negligible on a small scale, where the interactions of individual charged particles are important. On a large scale, such as between the Earth and a person, the reverse is true. Most objects are nearly electrically neutral, and so attractive and repulsive **Coulomb forces** nearly cancel. Gravitational force on a large scale dominates interactions between large objects because it is always attractive, while Coulomb forces tend to cancel.

Section Summary

- Frenchman Charles Coulomb was the first to publish the mathematical equation that describes the electrostatic force between two objects.
- Coulomb's law gives the magnitude of the force between point charges. It is
  Equation:

$$F = k\frac{|q_1 q_2|}{r^2},$$

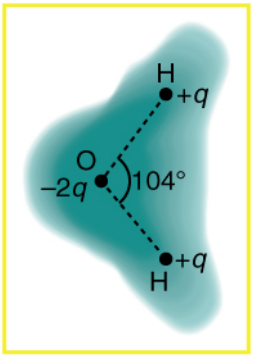
  where q_1 and q_2 are two point charges separated by a distance r, and $k \approx 8.99 \times 10^9\ \mathrm{N \cdot m^2/C^2}$
- This Coulomb force is extremely basic, since most charges are due to point-like particles. It is responsible for all electrostatic effects and underlies most macroscopic forces.
- The Coulomb force is extraordinarily strong compared with the gravitational force, another basic force—but unlike gravitational force it can cancel, since it can be either attractive or repulsive.
- The electrostatic force between two subatomic particles is far greater than the gravitational force between the same two particles.

Conceptual Questions

Exercise:

Problem:

[link] shows the charge distribution in a water molecule, which is called a polar molecule because it has an inherent separation of charge. Given water's polar character, explain what effect humidity has on removing excess charge from objects.



Schematic representation of the outer electron cloud of a neutral water molecule. The electrons spend more time near the oxygen than the hydrogens, giving a permanent charge separation as shown. Water is thus a *polar molecule*. It is more easily affected by electrostatic forces than molecules with uniform charge distributions.

Exercise:

Problem:

Using [link], explain, in terms of Coulomb's law, why a polar molecule (such as in [link]) is attracted by both positive and negative charges.

Exercise:

Problem:

Given the polar character of water molecules, explain how ions in the air form nucleation centers for rain droplets.

Problems & Exercises

Exercise:

Problem:

What is the repulsive force between two pith balls that are 8.00 cm apart and have equal charges of – 30.0 nC?

Exercise:

Problem:

(a) How strong is the attractive force between a glass rod with a $0.700\ \mu\text{C}$ charge and a silk cloth with a $-0.600\ \mu\text{C}$ charge, which are 12.0 cm apart, using the approximation that they act like point charges? (b) Discuss how the answer to this problem might be affected if the charges are distributed over some area and do not act like point charges.

Solution:

(a) 0.263 N

(b) If the charges are distributed over some area, there will be a concentration of charge along the side closest to the oppositely charged object. This effect will increase the net force.

Exercise:

Problem:

Two point charges exert a 5.00 N force on each other. What will the force become if the distance between them is increased by a factor of three?

Exercise:

Problem:

Two point charges are brought closer together, increasing the force between them by a factor of 25. By what factor was their separation decreased?

Solution:

The separation decreased by a factor of 5.

Exercise:

Problem:

How far apart must two point charges of 75.0 nC (typical of static electricity) be to have a force of 1.00 N between them?

Exercise:

Problem:

If two equal charges each of 1 C each are separated in air by a distance of 1 km, what is the magnitude of the force acting between them? You will see that even at a distance as large as 1 km, the repulsive force is substantial because 1 C is a very significant amount of charge.

Exercise:

Problem:

A test charge of $+2\ \mu\text{C}$ is placed halfway between a charge of $+6\ \mu\text{C}$ and another of $+4\ \mu\text{C}$ separated by 10 cm. (a) What is the magnitude of the force on the test charge? (b) What is the direction of this force (away from or toward the $+6\ \mu\text{C}$ charge)?

Exercise:

Problem:

Bare free charges do not remain stationary when close together. To illustrate this, calculate the acceleration of two isolated protons separated by 2.00 nm (a typical distance between gas atoms). Explicitly show how you follow the steps in the Problem-Solving Strategy for electrostatics.

Solution:

$$\begin{aligned} F &= k\frac{|q_1 q_2|}{r^2} = ma \Rightarrow a = \frac{kq^2}{mr^2} \\ &= \frac{\left(9.00\times10^{9}\ \text{N}\cdot\text{m}^2/\text{C}^2\right)\left(1.60\times10^{-19}\ \text{m}\right)^2}{\left(1.67\times10^{-27}\ \text{kg}\right)\left(2.00\times10^{-9}\ \text{m}\right)^2} \\ &= 3.45\times10^{16}\ \text{m/s}^2 \end{aligned}$$

Exercise:

Problem:

(a) By what factor must you change the distance between two point charges to change the force between them by a factor of 10? (b) Explain how the distance can either increase or decrease by this factor and still cause a factor of 10 change in the force.

Solution:

(a) 3.2

(b) If the distance increases by 3.2, then the force will decrease by a factor of 10 ; if the distance decreases by 3.2, then the force will increase by a factor of 10. Either way, the force changes by a factor of 10.

Exercise:

Problem:

Suppose you have a total charge q_{tot} that you can split in any manner. Once split, the separation distance is fixed. How do you split the charge to achieve the greatest force?

Exercise:

Problem:

(a) Common transparent tape becomes charged when pulled from a dispenser. If one piece is placed above another, the repulsive force can be great enough to support the top piece's weight. Assuming equal point charges (only an approximation), calculate the magnitude of the charge if electrostatic force is great enough to support the weight of a 10.0 mg piece of tape held 1.00 cm above another. (b) Discuss whether the magnitude of this charge is consistent with what is typical of static electricity.

Solution:

(a) 1.04×10^{-9} C

(b) This charge is approximately 1 nC, which is consistent with the magnitude of charge typical for static electricity

Exercise:

Problem:

(a) Find the ratio of the electrostatic to gravitational force between two electrons. (b) What is this ratio for two protons? (c) Why is the ratio different for electrons and protons?

Exercise:

Problem:

At what distance is the electrostatic force between two protons equal to the weight of one proton?

Exercise:

Problem:

A certain five cent coin contains 5.00 g of nickel. What fraction of the nickel atoms' electrons, removed and placed 1.00 m above it, would support the weight of this coin? The atomic mass of nickel is 58.7, and each nickel atom contains 28 electrons and 28 protons.

Solution:

1.02×10^{-11}

Exercise:

Problem:

(a) Two point charges totaling $8.00\ \mu\text{C}$ exert a repulsive force of 0.150 N on one another when separated by 0.500 m. What is the charge on each? (b) What is the charge on each if the force is attractive?

Exercise:

Problem:

Point charges of $5.00\ \mu\text{C}$ and $-3.00\ \mu\text{C}$ are placed 0.250 m apart. (a) Where can a third charge be placed so that the net force on it is zero? (b) What if both charges are positive?

Solution:

a. 0.859 m beyond negative charge on line connecting two charges
b. 0.109 m from lesser charge on line connecting two charges

Exercise:

Problem:

Two point charges q_1 and q_2 are $3.00\ \text{m}$ apart, and their total charge is $20\ \mu\text{C}$. (a) If the force of repulsion between them is 0.075N, what are magnitudes of the two charges? (b) If one charge attracts the other with a force of 0.525N, what are the magnitudes of the two charges? Note that you may need to solve a quadratic equation to reach your answer.

Glossary

Coulomb's law
: the mathematical equation calculating the electrostatic force vector between two charged particles

Coulomb force
: another term for the electrostatic force

electrostatic force
: the amount and direction of attraction or repulsion between two charged bodies

Electric Field: Concept of a Field Revisited

- Describe a force field and calculate the strength of an electric field due to a point charge.
- Calculate the force exerted on a test charge by an electric field.
- Explain the relationship between electrical force (F) on a test charge and electrical field strength (E).

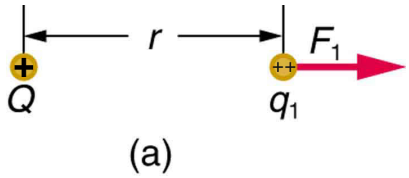
Contact forces, such as between a baseball and a bat, are explained on the small scale by the interaction of the charges in atoms and molecules in close proximity. They interact through forces that include the **Coulomb force**. Action at a distance is a force between objects that are not close enough for their atoms to "touch." That is, they are separated by more than a few atomic diameters.

For example, a charged rubber comb attracts neutral bits of paper from a distance via the Coulomb force. It is very useful to think of an object being surrounded in space by a **force field**. The force field carries the force to another object (called a test object) some distance away.

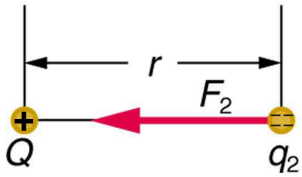
Concept of a Field

A field is a way of conceptualizing and mapping the force that surrounds any object and acts on another object at a distance without apparent physical connection. For example, the gravitational field surrounding the earth (and all other masses) represents the gravitational force that would be experienced if another mass were placed at a given point within the field.

In the same way, the Coulomb force field surrounding any charge extends throughout space. Using Coulomb's law, $F = k|q_1 q_2|/r^2$, its magnitude is given by the equation $F = k|\,qQ\,|/r^2$, for a **point charge** (a particle having a charge Q) acting on a **test charge** q at a distance r (see [link]). Both the magnitude and direction of the Coulomb force field depend on Q and the test charge q.



(a)



(b)

The Coulomb force field due to a positive charge Q is shown acting on two different charges. Both charges are the same distance from Q. (a) Since q_1 is positive, the force F_1 acting on it is repulsive. (b) The charge q_2 is negative and greater in magnitude than q_1, and so the force F_2 acting on it is attractive and stronger than F_1. The Coulomb force field is thus not unique at any point in space, because it depends on the test charges q_1 and q_2

as well as the
charge Q.

To simplify things, we would prefer to have a field that depends only on Q and not on the test charge *q*. The electric field is defined in such a manner that it represents only the charge creating it and is unique at every point in space. Specifically, the electric field E is defined to be the ratio of the Coulomb force to the test charge:
Equation:

$$\mathbf{E} = \frac{\mathbf{F}}{q},$$

where $\mathbf{F}$ is the electrostatic force (or Coulomb force) exerted on a positive test charge *q*. It is understood that $\mathbf{E}$ is in the same direction as $\mathbf{F}$. It is also assumed that *q* is so small that it does not alter the charge distribution creating the electric field. The units of electric field are newtons per coulomb (N/C). If the electric field is known, then the electrostatic force on any charge *q* is simply obtained by multiplying charge times electric field, or $\mathbf{F} = q\mathbf{E}$. Consider the electric field due to a point charge Q. According to Coulomb's law, the force it exerts on a test charge *q* is $F = k|\,qQ\,|/r^2$. Thus the magnitude of the electric field, E, for a point charge is
Equation:

$$E = \left|\frac{F}{q}\right| = k\left|\frac{\mathrm{qQ}}{qr^2}\right| = k\frac{|Q|}{r^2}.$$

Since the test charge cancels, we see that
Equation:

$$E = k\frac{|Q|}{r^2}.$$

The electric field is thus seen to depend only on the charge Q and the distance r; it is completely independent of the test charge q.

Example:
Calculating the Electric Field of a Point Charge
Calculate the strength and direction of the electric field E due to a point charge of 2.00 nC (nano-Coulombs) at a distance of 5.00 mm from the charge.
Strategy
We can find the electric field created by a point charge by using the equation $E = \mathrm{kQ}/r^2$.
Solution
Here $Q = 2.00 \times 10^{-9}$ C and $r = 5.00 \times 10^{-3}$ m. Entering those values into the above equation gives
Equation:

$$\begin{aligned} E &= k\frac{Q}{r^2} \\ &= (8.99 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2) \times \frac{(2.00 \times 10^{-9} \text{ C})}{(5.00 \times 10^{-3} \text{ m})^2} \\ &= 7.19 \times 10^5 \text{ N/C}. \end{aligned}$$

Discussion
This **electric field strength** is the same at any point 5.00 mm away from the charge Q that creates the field. It is positive, meaning that it has a direction pointing away from the charge Q.

Example:
Calculating the Force Exerted on a Point Charge by an Electric Field
What force does the electric field found in the previous example exert on a point charge of –0.250 μC?
Strategy
Since we know the electric field strength and the charge in the field, the force on that charge can be calculated using the definition of electric field

$\mathbf{E} = \mathbf{F}/q$ rearranged to $\mathbf{F} = q\mathbf{E}$.

Solution

The magnitude of the force on a charge $q = -0.250\ \mu\text{C}$ exerted by a field of strength $E = 7.20 \times 10^5$ N/C is thus,

Equation:

$$\begin{aligned} F &= -qE \\ &= (0.250 \times 10^{-6}\ \text{C})(7.20 \times 10^5\ \text{N/C}) \\ &= 0.180\ \text{N}. \end{aligned}$$

Because q is negative, the force is directed opposite to the direction of the field.

Discussion

The force is attractive, as expected for unlike charges. (The field was created by a positive charge and here acts on a negative charge.) The charges in this example are typical of common static electricity, and the modest attractive force obtained is similar to forces experienced in static cling and similar situations.

Note:

PhET Explorations: Electric Field of Dreams

Play ball! Add charges to the Field of Dreams and see how they react to the electric field. Turn on a background electric field and adjust the direction and magnitude.

https://archive.cnx.org/specials/ca9a78b4-06a7-11e6-b638-3bb71d1f0b42/electric-field-of-dreams/#sim-electric-field-of-dreams

Section Summary

- The electrostatic force field surrounding a charged object extends out into space in all directions.
- The electrostatic force exerted by a point charge on a test charge at a distance r depends on the charge of both charges, as well as the

distance between the two.

- The electric field $\mathbf{E}$ is defined to be
  Equation:

$$\mathbf{E} = \frac{\mathbf{F}}{q},$$

  where $\mathbf{F}$ is the Coulomb or electrostatic force exerted on a small positive test charge q. $\mathbf{E}$ has units of N/C.
- The magnitude of the electric field $\mathbf{E}$ created by a point charge Q is
  Equation:

$$\mathbf{E} = k\frac{|Q|}{r^2}.$$

  where r is the distance from Q. The electric field $\mathbf{E}$ is a vector and fields due to multiple charges add like vectors.

Conceptual Questions

Exercise:

Problem:

Why must the test charge q in the definition of the electric field be vanishingly small?

Exercise:

Problem:

Are the direction and magnitude of the Coulomb force unique at a given point in space? What about the electric field?

Problem Exercises

Exercise:

Problem:

What is the magnitude and direction of an electric field that exerts a 2.00×10^{-5} N upward force on a –1.75 μC charge?

Exercise:

Problem:

What is the magnitude and direction of the force exerted on a $3.50\ \mu\text{C}$ charge by a 250 N/C electric field that points due east?

Solution:

8.75×10^{-4} N

Exercise:

Problem:

Calculate the magnitude of the electric field 2.00 m from a point charge of 5.00 mC (such as found on the terminal of a Van de Graaff).

Exercise:

Problem:

(a) What magnitude point charge creates a 10,000 N/C electric field at a distance of 0.250 m? (b) How large is the field at 10.0 m?

Solution:

(a) 6.94×10^{-8} C

(b) $6.25\ \text{N/C}$

Exercise:

Problem:

Calculate the initial (from rest) acceleration of a proton in a $5.00 \times 10^6\ \mathrm{N/C}$ electric field (such as created by a research Van de Graaff). Explicitly show how you follow the steps in the Problem-Solving Strategy for electrostatics.

Exercise:

Problem:

(a) Find the magnitude and direction of an electric field that exerts a $4.80 \times 10^{-17}\ \mathrm{N}$ westward force on an electron. (b) What magnitude and direction force does this field exert on a proton?

Solution:

(a) $300\ \mathrm{N/C}\ (\mathrm{east})$

(b) $4.80 \times 10^{-17}\ \mathrm{N}\ (\mathrm{east})$

Glossary

field
: a map of the amount and direction of a force acting on other objects, extending out into space

point charge
: A charged particle, designated Q, generating an electric field

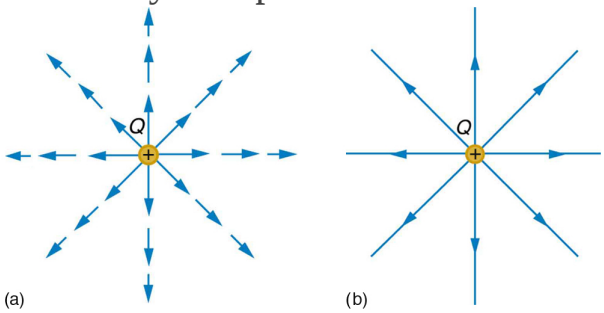
test charge
: A particle (designated *q*) with either a positive or negative charge set down within an electric field generated by a point charge

Electric Field Lines: Multiple Charges

- Calculate the total force (magnitude and direction) exerted on a test charge from more than one charge
- Describe an electric field diagram of a positive point charge; of a negative point charge with twice the magnitude of positive charge
- Draw the electric field lines between two points of the same charge; between two points of opposite charge.

Drawings using lines to represent **electric fields** around charged objects are very useful in visualizing field strength and direction. Since the electric field has both magnitude and direction, it is a vector. Like all **vectors**, the electric field can be represented by an arrow that has length proportional to its magnitude and that points in the correct direction. (We have used arrows extensively to represent force vectors, for example.)

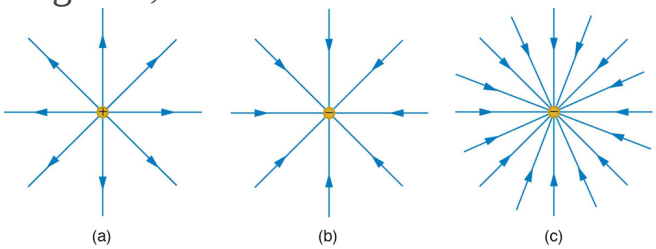
[link] shows two pictorial representations of the same electric field created by a positive point charge Q. [link] (b) shows the standard representation using continuous lines. [link] (a) shows numerous individual arrows with each arrow representing the force on a test charge q. Field lines are essentially a map of infinitesimal force vectors.



Two equivalent representations of the electric field due to a positive charge Q. (a) Arrows representing the electric field's magnitude and direction. (b) In the standard representation, the arrows are replaced by continuous field lines having the same direction at any point

as the electric field. The closeness of the lines is directly related to the strength of the electric field. A test charge placed anywhere will feel a force in the direction of the field line; this force will have a strength proportional to the density of the lines (being greater near the charge, for example).

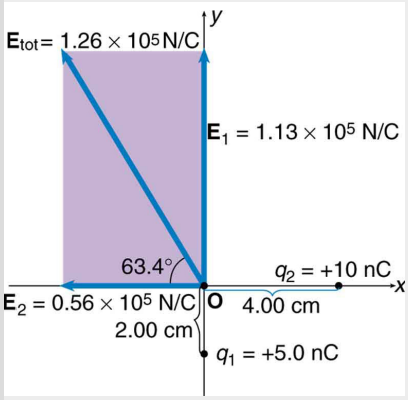
Note that the electric field is defined for a positive test charge q, so that the field lines point away from a positive charge and toward a negative charge. (See [link].) The electric field strength is exactly proportional to the number of field lines per unit area, since the magnitude of the electric field for a point charge is $E = k|Q|/r^2$ and area is proportional to r^2. This pictorial representation, in which field lines represent the direction and their closeness (that is, their areal density or the number of lines crossing a unit area) represents strength, is used for all fields: electrostatic, gravitational, magnetic, and others.



The electric field surrounding three different point charges. (a) A positive charge. (b) A negative charge of equal magnitude. (c) A larger negative charge.

In many situations, there are multiple charges. The total electric field created by multiple charges is the vector sum of the individual fields created by each charge. The following example shows how to add electric field vectors.

Example:
Adding Electric Fields
Find the magnitude and direction of the total electric field due to the two point charges, q_1 and q_2, at the origin of the coordinate system as shown in [link].



The electric fields $\mathbf{E}_1$ and $\mathbf{E}_2$ at the origin O add to $\mathbf{E}_{\text{tot}}$.

Strategy
Since the electric field is a vector (having magnitude and direction), we add electric fields with the same vector techniques used for other types of vectors. We first must find the electric field due to each charge at the point of interest, which is the origin of the coordinate system (O) in this instance. We pretend that there is a positive test charge, q, at point O, which allows us to determine the direction of the fields $\mathbf{E}_1$ and $\mathbf{E}_2$. Once those fields are found, the total field can be determined using **vector addition**.
Solution

The electric field strength at the origin due to q_1 is labeled E_1 and is calculated:
Equation:

$$E_1 = k\frac{q_1}{r_1^2} = \left(8.99 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2\right) \frac{\left(5.00 \times 10^{-9} \text{ C}\right)}{\left(2.00 \times 10^{-2} \text{ m}\right)^2}$$
$$E_1 = 1.124 \times 10^5 \text{ N/C}.$$

Similarly, E_2 is
Equation:

$$E_2 = k\frac{q_2}{r_2^2} = \left(8.99 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2\right) \frac{\left(10.0 \times 10^{-9} \text{ C}\right)}{\left(4.00 \times 10^{-2} \text{ m}\right)^2}$$
$$E_2 = 0.5619 \times 10^5 \text{ N/C}.$$

Four digits have been retained in this solution to illustrate that E_1 is exactly twice the magnitude of E_2. Now arrows are drawn to represent the magnitudes and directions of $\mathbf{E}_1$ and $\mathbf{E}_2$. (See [link].) The direction of the electric field is that of the force on a positive charge so both arrows point directly away from the positive charges that create them. The arrow for $\mathbf{E}_1$ is exactly twice the length of that for $\mathbf{E}_2$. The arrows form a right triangle in this case and can be added using the Pythagorean theorem. The magnitude of the total field E_{tot} is
Equation:

$$\begin{aligned} E_{\text{tot}} &= (E_1^2 + E_2^2)^{1/2} \\ &= \{(1.124 \times 10^5 \text{ N/C})^2 + (0.5619 \times 10^5 \text{ N/C})^2\}^{1/2} \\ &= 1.26 \times 10^5 \text{ N/C}. \end{aligned}$$

The direction is
Equation:

$$\begin{aligned}\theta &= \tan^{-1}\left(\frac{E_1}{E_2}\right)\\ &= \tan^{-1}\left(\frac{1.124\times10^5\ \text{N/C}}{0.5619\times10^5\ \text{N/C}}\right)\\ &= 63.4^\text{o},\end{aligned}$$

or 63.4^o above the *x*-axis.

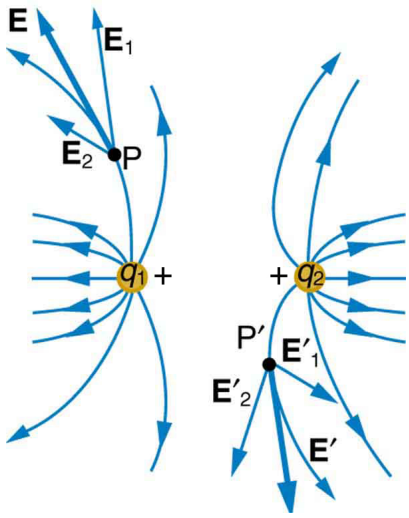
Discussion

In cases where the electric field vectors to be added are not perpendicular, vector components or graphical techniques can be used. The total electric field found in this example is the total electric field at only one point in space. To find the total electric field due to these two charges over an entire region, the same technique must be repeated for each point in the region. This impossibly lengthy task (there are an infinite number of points in space) can be avoided by calculating the total field at representative points and using some of the unifying features noted next.

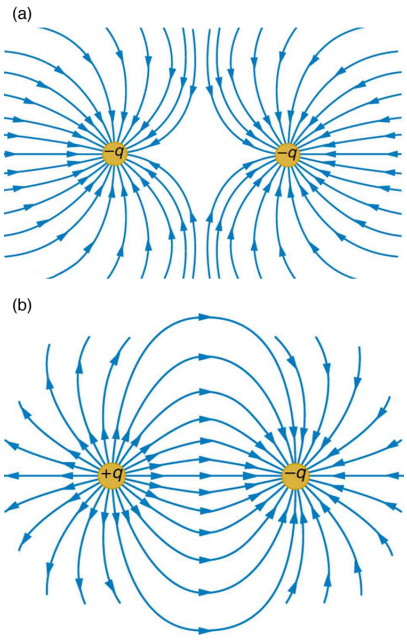
[link] shows how the electric field from two point charges can be drawn by finding the total field at representative points and drawing electric field lines consistent with those points. While the electric fields from multiple charges are more complex than those of single charges, some simple features are easily noticed.

For example, the field is weaker between like charges, as shown by the lines being farther apart in that region. (This is because the fields from each charge exert opposing forces on any charge placed between them.) (See [link] and [link](a).) Furthermore, at a great distance from two like charges, the field becomes identical to the field from a single, larger charge.

[link](b) shows the electric field of two unlike charges. The field is stronger between the charges. In that region, the fields from each charge are in the same direction, and so their strengths add. The field of two unlike charges is weak at large distances, because the fields of the individual charges are in opposite directions and so their strengths subtract. At very large distances, the field of two unlike charges looks like that of a smaller single charge.



Two positive point charges q_1 and q_2 produce the resultant electric field shown. The field is calculated at representative points and then smooth field lines drawn following the rules outlined in the text.



(a) Two negative charges produce the fields shown. It is very similar to the field produced by two positive charges, except that the directions are reversed. The field is clearly weaker between the charges. The individual forces on a test charge in that region are in opposite directions. (b) Two opposite charges produce the field shown, which is stronger in the region between the charges.

We use electric field lines to visualize and analyze electric fields (the lines are a pictorial tool, not a physical entity in themselves). The properties of electric field lines for any charge distribution can be summarized as follows:

1. Field lines must begin on positive charges and terminate on negative charges, or at infinity in the hypothetical case of isolated charges.
2. The number of field lines leaving a positive charge or entering a negative charge is proportional to the magnitude of the charge.
3. The strength of the field is proportional to the closeness of the field lines—more precisely, it is proportional to the number of lines per unit area perpendicular to the lines.
4. The direction of the electric field is tangent to the field line at any point in space.
5. Field lines can never cross.

The last property means that the field is unique at any point. The field line represents the direction of the field; so if they crossed, the field would have two directions at that location (an impossibility if the field is unique).

Note:
PhET Explorations: Charges and Fields
Move point charges around on the playing field and then view the electric field, voltages, equipotential lines, and more. It's colorful, it's dynamic, it's free.

Click here for the simulation .

Section Summary

- Drawings of electric field lines are useful visual tools. The properties of electric field lines for any charge distribution are that:
- Field lines must begin on positive charges and terminate on negative charges, or at infinity in the hypothetical case of isolated charges.
- The number of field lines leaving a positive charge or entering a negative charge is proportional to the magnitude of the charge.
- The strength of the field is proportional to the closeness of the field lines—more precisely, it is proportional to the number of lines per unit area perpendicular to the lines.
- The direction of the electric field is tangent to the field line at any point in space.
- Field lines can never cross.

Conceptual Questions

Exercise:

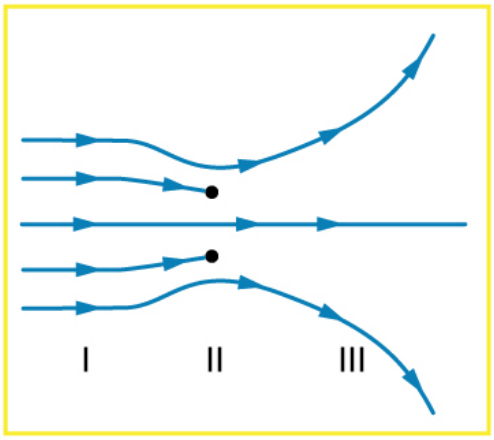
Problem:

Compare and contrast the Coulomb force field and the electric field. To do this, make a list of five properties for the Coulomb force field analogous to the five properties listed for electric field lines. Compare each item in your list of Coulomb force field properties with those of the electric field—are they the same or different? (For example, electric field lines cannot cross. Is the same true for Coulomb field lines?)

Exercise:

Problem:

[link] shows an electric field extending over three regions, labeled I, II, and III. Answer the following questions. (a) Are there any isolated charges? If so, in what region and what are their signs? (b) Where is the field strongest? (c) Where is it weakest? (d) Where is the field the most uniform?



Problem Exercises

Exercise:

Problem:

(a) Sketch the electric field lines near a point charge $+q$. (b) Do the same for a point charge $-3.00q$.

Exercise:

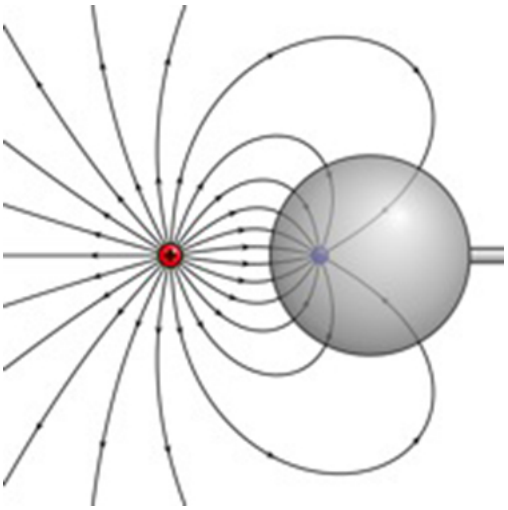
Problem:

Sketch the electric field lines a long distance from the charge distributions shown in [link] (a) and (b)

Exercise:

Problem:

[link] shows the electric field lines near two charges q_1 and q_2. What is the ratio of their magnitudes? (b) Sketch the electric field lines a long distance from the charges shown in the figure.



The electric field near two charges.

Exercise:

Problem:

Sketch the electric field lines in the vicinity of two opposite charges, where the negative charge is three times greater in magnitude than the positive. (See [link] for a similar situation).

Glossary

electric field
: a three-dimensional map of the electric force extended out into space from a point charge

electric field lines
: a series of lines drawn from a point charge representing the magnitude and direction of force exerted by that charge

vector
: a quantity with both magnitude and direction

vector addition

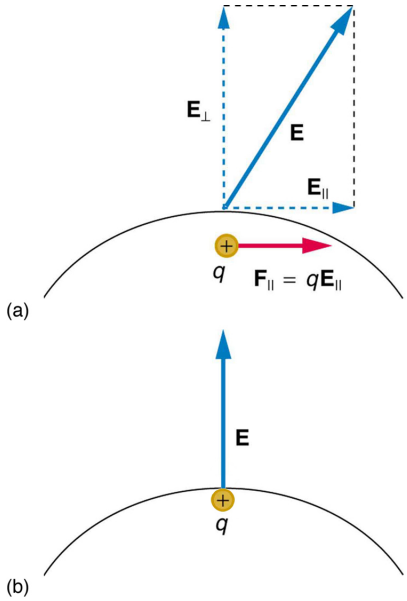
: mathematical combination of two or more vectors, including their magnitudes, directions, and positions

Conductors and Electric Fields in Static Equilibrium

- List the three properties of a conductor in electrostatic equilibrium.
- Explain the effect of an electric field on free charges in a conductor.
- Explain why no electric field may exist inside a conductor.
- Describe the electric field surrounding Earth.
- Explain what happens to an electric field applied to an irregular conductor.
- Describe how a lightning rod works.
- Explain how a metal car may protect passengers inside from the dangerous electric fields caused by a downed line touching the car.

Conductors contain **free charges** that move easily. When excess charge is placed on a conductor or the conductor is put into a static electric field, charges in the conductor quickly respond to reach a steady state called **electrostatic equilibrium**.

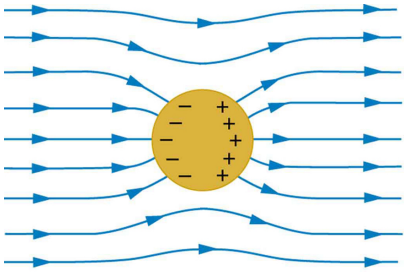
[link] shows the effect of an electric field on free charges in a conductor. The free charges move until the field is perpendicular to the conductor's surface. There can be no component of the field parallel to the surface in electrostatic equilibrium, since, if there were, it would produce further movement of charge. A positive free charge is shown, but free charges can be either positive or negative and are, in fact, negative in metals. The motion of a positive charge is equivalent to the motion of a negative charge in the opposite direction.



When an electric field $\mathbf{E}$ is applied to a conductor, free charges inside the conductor move until the field is perpendicular to the surface. (a) The electric field is a vector quantity, with both parallel and perpendicular components. The parallel component ($\mathbf{E}_{||}$) exerts a force ($\mathbf{F}_{||}$) on the free charge q, which moves the charge until $\mathbf{F}_{||} = 0$. (b) The resulting field is perpendicular to the surface. The free charge has

been brought to the conductor’s surface, leaving electrostatic forces in equilibrium.

A conductor placed in an **electric field** will be **polarized**. [link] shows the result of placing a neutral conductor in an originally uniform electric field. The field becomes stronger near the conductor but entirely disappears inside it.



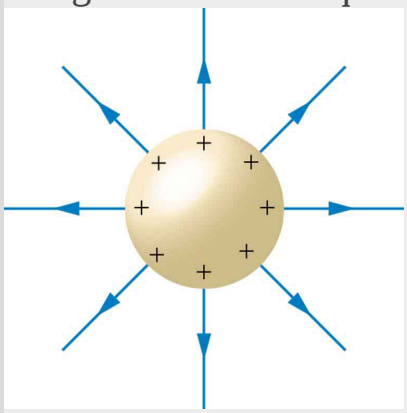
This illustration shows a spherical conductor in static equilibrium with an originally uniform electric field. Free charges move within the conductor, polarizing it, until the electric field lines are perpendicular to the surface. The field lines end on excess negative charge on one section of the surface and begin

again on excess positive charge on the opposite side. No electric field exists inside the conductor, since free charges in the conductor would continue moving in response to any field until it was neutralized.

Note:

Misconception Alert: Electric Field inside a Conductor

Excess charges placed on a spherical conductor repel and move until they are evenly distributed, as shown in [link]. Excess charge is forced to the surface until the field inside the conductor is zero. Outside the conductor, the field is exactly the same as if the conductor were replaced by a point charge at its center equal to the excess charge.



The mutual repulsion of excess positive charges on

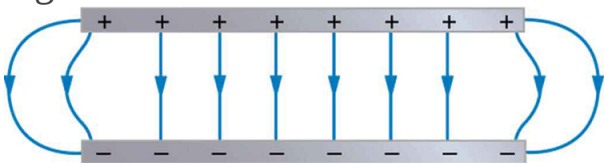
a spherical conductor distributes them uniformly on its surface. The resulting electric field is perpendicular to the surface and zero inside. Outside the conductor, the field is identical to that of a point charge at the center equal to the excess charge.

Note:
Properties of a Conductor in Electrostatic Equilibrium

1. The electric field is zero inside a conductor.
2. Just outside a conductor, the electric field lines are perpendicular to its surface, ending or beginning on charges on the surface.
3. Any excess charge resides entirely on the surface or surfaces of a conductor.

The properties of a conductor are consistent with the situations already discussed and can be used to analyze any conductor in electrostatic equilibrium. This can lead to some interesting new insights, such as described below.

How can a very uniform electric field be created? Consider a system of two metal plates with opposite charges on them, as shown in [link]. The properties of conductors in electrostatic equilibrium indicate that the electric field between the plates will be uniform in strength and direction. Except near the edges, the excess charges distribute themselves uniformly, producing field lines that are uniformly spaced (hence uniform in strength) and perpendicular to the surfaces (hence uniform in direction, since the plates are flat). The edge effects are less important when the plates are close together.



Two metal plates with equal, but opposite, excess charges. The field between them is uniform in strength and direction except near the edges. One use of such a field is to produce uniform acceleration of charges between the plates, such as in the electron gun of a TV tube.

Earth's Electric Field

A near uniform electric field of approximately 150 N/C, directed downward, surrounds Earth, with the magnitude increasing slightly as we get closer to the surface. What causes the electric field? At around 100 km above the surface of Earth we have a layer of charged particles, called the **ionosphere**. The ionosphere is responsible for a range of phenomena including the electric field surrounding Earth. In fair weather the ionosphere is positive and the Earth largely negative, maintaining the electric field ([link](a)).

In storm conditions clouds form and localized electric fields can be larger and reversed in direction ([link](b)). The exact charge distributions depend on the local conditions, and variations of [link](b) are possible.

If the electric field is sufficiently large, the insulating properties of the surrounding material break down and it becomes conducting. For air this occurs at around 3×10^6 N/C. Air ionizes ions and electrons recombine, and we get discharge in the form of lightning sparks and corona discharge.



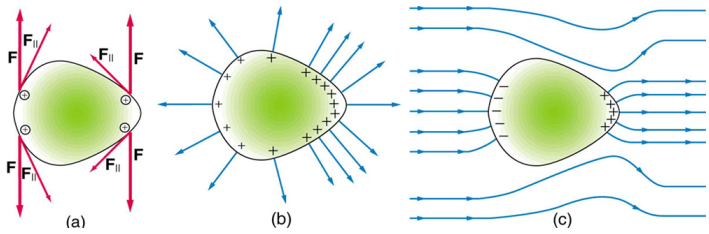
Earth’s electric field. (a) Fair weather field. Earth and the ionosphere (a layer of charged particles) are both conductors. They produce a uniform electric field of about 150 N/C. (credit: D. H. Parks) (b) Storm fields. In the presence of storm clouds, the local electric fields can be larger. At very high fields, the insulating properties of the air break down and lightning can occur. (credit: Jan-Joost Verhoef)

Electric Fields on Uneven Surfaces

So far we have considered excess charges on a smooth, symmetrical conductor surface. What happens if a conductor has sharp corners or is pointed? Excess charges on a nonuniform conductor become concentrated at the sharpest points. Additionally, excess charge may move on or off the conductor at the sharpest points.

To see how and why this happens, consider the charged conductor in [link]. The electrostatic repulsion of like charges is most effective in moving them apart on the flattest surface, and so they become least concentrated there. This is because the forces between identical pairs of charges at either end of the conductor are identical, but the components of the forces parallel to the surfaces are different. The component parallel to the surface is greatest on the flattest surface and, hence, more effective in moving the charge.

The same effect is produced on a conductor by an externally applied electric field, as seen in [link] (c). Since the field lines must be perpendicular to the surface, more of them are concentrated on the most curved parts.



Excess charge on a nonuniform conductor becomes most concentrated at the location of greatest curvature. (a) The forces between identical pairs of charges at either end of the conductor are identical, but the components of the forces parallel to the surface are different. It is $\mathbf{F}_{||}$ that moves the charges apart once they

have reached the surface. (b) $\mathbf{F}_{\|}$ is smallest at the more pointed end, the charges are left closer together, producing the electric field shown. (c) An uncharged conductor in an originally uniform electric field is polarized, with the most concentrated charge at its most pointed end.

Applications of Conductors

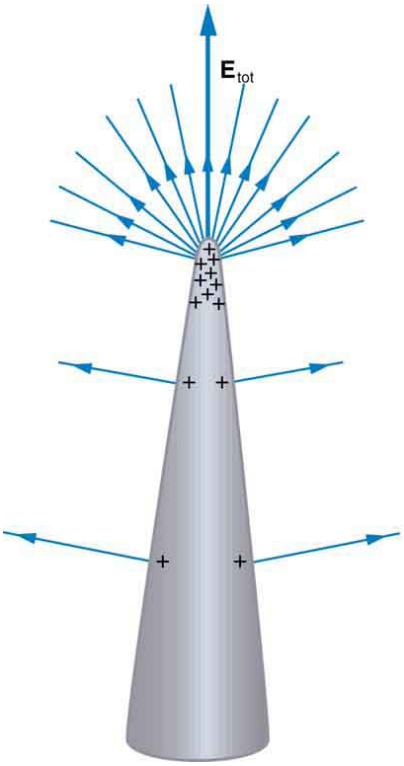
On a very sharply curved surface, such as shown in [link], the charges are so concentrated at the point that the resulting electric field can be great enough to remove them from the surface. This can be useful.

Lightning rods work best when they are most pointed. The large charges created in storm clouds induce an opposite charge on a building that can result in a lightning bolt hitting the building. The induced charge is bled away continually by a lightning rod, preventing the more dramatic lightning strike.

Of course, we sometimes wish to prevent the transfer of charge rather than to facilitate it. In that case, the conductor should be very smooth and have as large a radius of curvature as possible. (See [link].) Smooth surfaces are used on high-voltage transmission lines, for example, to avoid leakage of charge into the air.

Another device that makes use of some of these principles is a **Faraday cage**. This is a metal shield that encloses a volume. All electrical charges will reside on the outside surface of this shield, and there will be no electrical field inside. A Faraday cage is used to prohibit stray electrical fields in the environment from interfering with sensitive measurements, such as the electrical signals inside a nerve cell.

During electrical storms if you are driving a car, it is best to stay inside the car as its metal body acts as a Faraday cage with zero electrical field inside. If in the vicinity of a lightning strike, its effect is felt on the outside of the car and the inside is unaffected, provided you remain totally inside. This is also true if an active (“hot”) electrical wire was broken (in a storm or an accident) and fell on your car.



A very pointed conductor has a large charge concentration at the point. The electric field is very strong at the point and can exert a force large enough to transfer charge on or off the conductor.

Lightning rods are used to prevent the buildup of large excess charges on structures and, thus, are pointed.



(a) (b)

(a) A lightning rod is pointed to facilitate the transfer of charge. (credit: Romaine, Wikimedia Commons) (b) This Van de Graaff generator has a smooth surface with a large radius of curvature to prevent the transfer of charge and allow a large voltage to be generated. The mutual repulsion of like charges is evident in the person’s hair while touching the metal sphere. (credit: Jon ‘ShakataGaNai’ Davis/Wikimedia Commons).

Section Summary

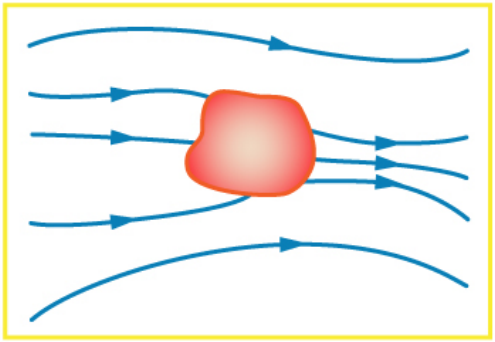
- A conductor allows free charges to move about within it.
- The electrical forces around a conductor will cause free charges to move around inside the conductor until static equilibrium is reached.
- Any excess charge will collect along the surface of a conductor.
- Conductors with sharp corners or points will collect more charge at those points.
- A lightning rod is a conductor with sharply pointed ends that collect excess charge on the building caused by an electrical storm and allow it to dissipate back into the air.
- Electrical storms result when the electrical field of Earth's surface in certain locations becomes more strongly charged, due to changes in the insulating effect of the air.
- A Faraday cage acts like a shield around an object, preventing electric charge from penetrating inside.

Conceptual Questions

Exercise:

Problem:

Is the object in [link] a conductor or an insulator? Justify your answer.



Exercise:

Problem:

If the electric field lines in the figure above were perpendicular to the object, would it necessarily be a conductor? Explain.

Exercise:

Problem:

The discussion of the electric field between two parallel conducting plates, in this module states that edge effects are less important if the plates are close together. What does close mean? That is, is the actual plate separation crucial, or is the ratio of plate separation to plate area crucial?

Exercise:

Problem:

Would the self-created electric field at the end of a pointed conductor, such as a lightning rod, remove positive or negative charge from the conductor? Would the same sign charge be removed from a neutral pointed conductor by the application of a similar externally created electric field? (The answers to both questions have implications for charge transfer utilizing points.)

Exercise:

Problem:

Why is a golfer with a metal club over her shoulder vulnerable to lightning in an open fairway? Would she be any safer under a tree?

Exercise:

Problem:

Can the belt of a Van de Graaff accelerator be a conductor? Explain.

Exercise:

Problem:

Are you relatively safe from lightning inside an automobile? Give two reasons.

Exercise:

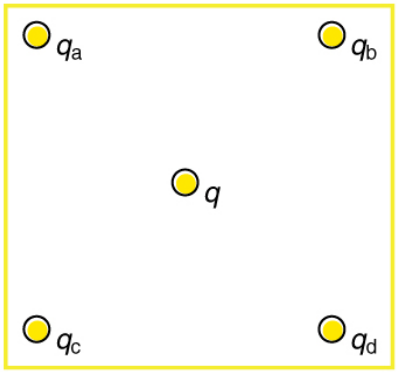
Problem:

Discuss pros and cons of a lightning rod being grounded versus simply being attached to a building.

Exercise:

Problem:

Using the symmetry of the arrangement, show that the net Coulomb force on the charge q at the center of the square below ([link]) is zero if the charges on the four corners are exactly equal.



Four point charges q_a, q_b, q_c, and q_d lie on the corners of a square and q is located at its center.

Exercise:

Problem:

(a) Using the symmetry of the arrangement, show that the electric field at the center of the square in [link] is zero if the charges on the four corners are exactly equal. (b) Show that this is also true for any combination of charges in which $q_a = q_d$ and $q_b = q_c$

Exercise:

Problem:

(a) What is the direction of the total Coulomb force on q in [link] if q is negative, $q_a = q_c$ and both are negative, and $q_b = q_c$ and both are positive? (b) What is the direction of the electric field at the center of the square in this situation?

Exercise:

Problem:

Considering [link], suppose that $q_a = q_d$ and $q_b = q_c$. First show that q is in static equilibrium. (You may neglect the gravitational force.) Then discuss whether the equilibrium is stable or unstable, noting that this may depend on the signs of the charges and the direction of displacement of q from the center of the square.

Exercise:

Problem:

If $q_a = 0$ in [link], under what conditions will there be no net Coulomb force on q?

Exercise:

Problem:

In regions of low humidity, one develops a special “grip” when opening car doors, or touching metal door knobs. This involves placing as much of the hand on the device as possible, not just the ends of one’s fingers. Discuss the induced charge and explain why this is done.

Exercise:

Problem:

Tollbooth stations on roadways and bridges usually have a piece of wire stuck in the pavement before them that will touch a car as it approaches. Why is this done?

Exercise:

Problem:

Suppose a woman carries an excess charge. To maintain her charged status can she be standing on ground wearing just any pair of shoes? How would you discharge her? What are the consequences if she simply walks away?

Problems & Exercises

Exercise:

Problem:

Sketch the electric field lines in the vicinity of the conductor in [link] given the field was originally uniform and parallel to the object's long axis. Is the resulting field small near the long side of the object?



Exercise:

Problem:

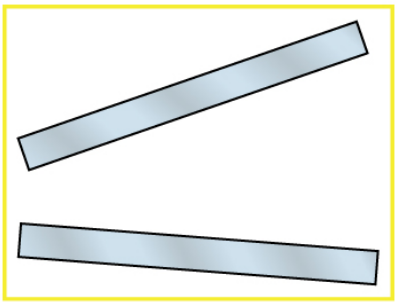
Sketch the electric field lines in the vicinity of the conductor in [link] given the field was originally uniform and parallel to the object's long axis. Is the resulting field small near the long side of the object?



Exercise:

Problem:

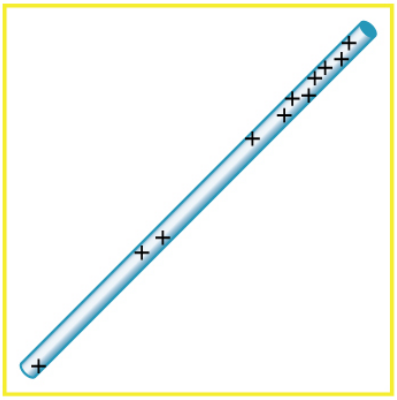
Sketch the electric field between the two conducting plates shown in [link], given the top plate is positive and an equal amount of negative charge is on the bottom plate. Be certain to indicate the distribution of charge on the plates.



Exercise:

Problem:

Sketch the electric field lines in the vicinity of the charged insulator in [link] noting its nonuniform charge distribution.



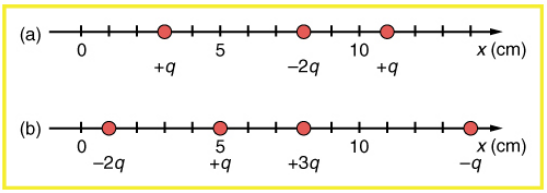
A charged insulating rod such as might be used in

a classroom demonstration.

Exercise:

Problem:

What is the force on the charge located at $x = 8.00$ cm in [link](a) given that $q = 1.00\ \mu\text{C}$?



(a) Point charges located at 3.00, 8.00, and 11.0 cm along the *x*-axis. (b) Point charges located at 1.00, 5.00, 8.00, and 14.0 cm along the *x*-axis.

Exercise:

Problem:

(a) Find the total electric field at $x = 1.00$ cm in [link](b) given that $q = 5.00\ \text{nC}$. (b) Find the total electric field at $x = 11.00$ cm in [link] (b). (c) If the charges are allowed to move and eventually be brought to rest by friction, what will the final charge configuration be? (That is, will there be a single charge, double charge, etc., and what will its value(s) be?)

Solution:

(a) $E_{x=1.00\ \text{cm}} = -\infty$

(b) $2.12 \times 10^5\ \mathrm{N/C}$

(c) one charge of $+q$

Exercise:

Problem:

(a) Find the electric field at $x = 5.00$ cm in [link](a), given that $q = 1.00\ \mu\mathrm{C}$. (b) At what position between 3.00 and 8.00 cm is the total electric field the same as that for $-2q$ alone? (c) Can the electric field be zero anywhere between 0.00 and 8.00 cm? (d) At very large positive or negative values of *x*, the electric field approaches zero in both (a) and (b). In which does it most rapidly approach zero and why? (e) At what position to the right of 11.0 cm is the total electric field zero, other than at infinity? (Hint: A graphing calculator can yield considerable insight in this problem.)

Exercise:

Problem:

(a) Find the total Coulomb force on a charge of 2.00 nC located at $x = 4.00$ cm in [link] (b), given that $q = 1.00\ \mu\mathrm{C}$. (b) Find the *x*-position at which the electric field is zero in [link] (b).

Solution:

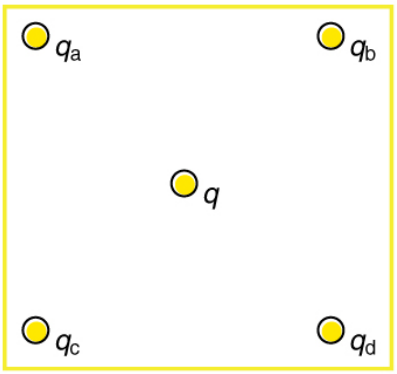
(a) 0.252 N to the left

(b) $x = 6.07$ cm

Exercise:

Problem:

Using the symmetry of the arrangement, determine the direction of the force on *q* in the figure below, given that $q_a = q_b = +7.50\ \mu\mathrm{C}$ and $q_c = q_d = -7.50\ \mu\mathrm{C}$. (b) Calculate the magnitude of the force on the charge *q*, given that the square is 10.0 cm on a side and $q = 2.00\ \mu\mathrm{C}$.



Exercise:

Problem:

(a) Using the symmetry of the arrangement, determine the direction of the electric field at the center of the square in [link], given that $q_a = q_b = -1.00\ \mu\text{C}$ and $q_c = q_d = +1.00\ \mu\text{C}$. (b) Calculate the magnitude of the electric field at the location of q, given that the square is 5.00 cm on a side.

Solution:

(a)The electric field at the center of the square will be straight up, since q_a and q_b are positive and q_c and q_d are negative and all have the same magnitude.

(b) $2.04 \times 10^7\ \text{N/C}\ (\text{upward})$

Exercise:

Problem:

Find the electric field at the location of q_a in [link] given that $q_b = q_c = q_d = +2.00\ \text{nC}$, $q = -1.00\ \text{nC}$, and the square is 20.0 cm on a side.

Exercise:

Problem:

Find the total Coulomb force on the charge q in [link], given that $q = 1.00\ \mu\text{C}$, $q_a = 2.00\ \mu\text{C}$, $q_b = -3.00\ \mu\text{C}$, $q_c = -4.00\ \mu\text{C}$, and $q_d = +1.00\ \mu\text{C}$. The square is 50.0 cm on a side.

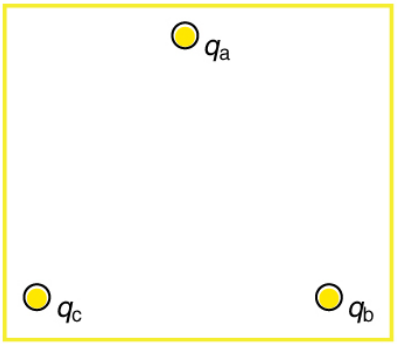
Solution:

$0.102\ \text{N}$, in the $-y$ direction

Exercise:

Problem:

(a) Find the electric field at the location of q_a in [link], given that $q_\text{b} = +10.00\ \mu\text{C}$ and $q_\text{c} = -5.00\ \mu\text{C}$. (b) What is the force on q_a, given that $q_\text{a} = +1.50\ \text{nC}$?



Point charges located at the corners of an equilateral triangle 25.0 cm on a side.

Exercise:

Problem:

(a) Find the electric field at the center of the triangular configuration of charges in [link], given that $q_a = +2.50\ \text{nC}$, $q_b = -8.00\ \text{nC}$, and $q_c = +1.50\ \text{nC}$. (b) Is there any combination of charges, other than $q_a = q_b = q_c$, that will produce a zero strength electric field at the center of the triangular configuration?

Solution:

(a) $\vec{E} = 4.36 \times 10^3\ \text{N/C}$, 35.0º, below the horizontal.

(b) No

Glossary

conductor
: an object with properties that allow charges to move about freely within it

free charge
: an electrical charge (either positive or negative) which can move about separately from its base molecule

electrostatic equilibrium
: an electrostatically balanced state in which all free electrical charges have stopped moving about

polarized
: a state in which the positive and negative charges within an object have collected in separate locations

ionosphere
: a layer of charged particles located around 100 km above the surface of Earth, which is responsible for a range of phenomena including the electric field surrounding Earth

Faraday cage
: a metal shield which prevents electric charge from penetrating its surface

Applications of Electrostatics

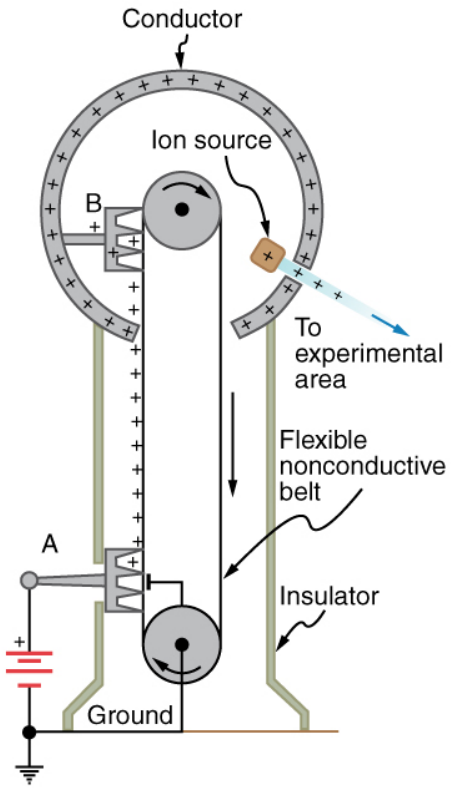
- Name several real-world applications of the study of electrostatics.

The study of **electrostatics** has proven useful in many areas. This module covers just a few of the many applications of electrostatics.

The Van de Graaff Generator

Van de Graaff generators (or Van de Graaffs) are not only spectacular devices used to demonstrate high voltage due to static electricity—they are also used for serious research. The first was built by Robert Van de Graaff in 1931 (based on original suggestions by Lord Kelvin) for use in nuclear physics research. [link] shows a schematic of a large research version. Van de Graaffs utilize both smooth and pointed surfaces, and conductors and insulators to generate large static charges and, hence, large voltages.

A very large excess charge can be deposited on the sphere, because it moves quickly to the outer surface. Practical limits arise because the large electric fields polarize and eventually ionize surrounding materials, creating free charges that neutralize excess charge or allow it to escape. Nevertheless, voltages of 15 million volts are well within practical limits.



Schematic of Van de Graaff generator. A battery (A) supplies excess positive charge to a pointed conductor, the points of which spray the charge onto a moving insulating belt near the bottom. The pointed conductor (B) on top in the large sphere picks up the charge. (The induced electric field at the points is so large that it removes the charge from the belt.) This can be done because the charge does not

remain inside the conducting sphere but moves to its outside surface. An ion source inside the sphere produces positive ions, which are accelerated away from the positive sphere to high velocities.

Note:
Take-Home Experiment: Electrostatics and Humidity
Rub a comb through your hair and use it to lift pieces of paper. It may help to tear the pieces of paper rather than cut them neatly. Repeat the exercise in your bathroom after you have had a long shower and the air in the bathroom is moist. Is it easier to get electrostatic effects in dry or moist air? Why would torn paper be more attractive to the comb than cut paper? Explain your observations.

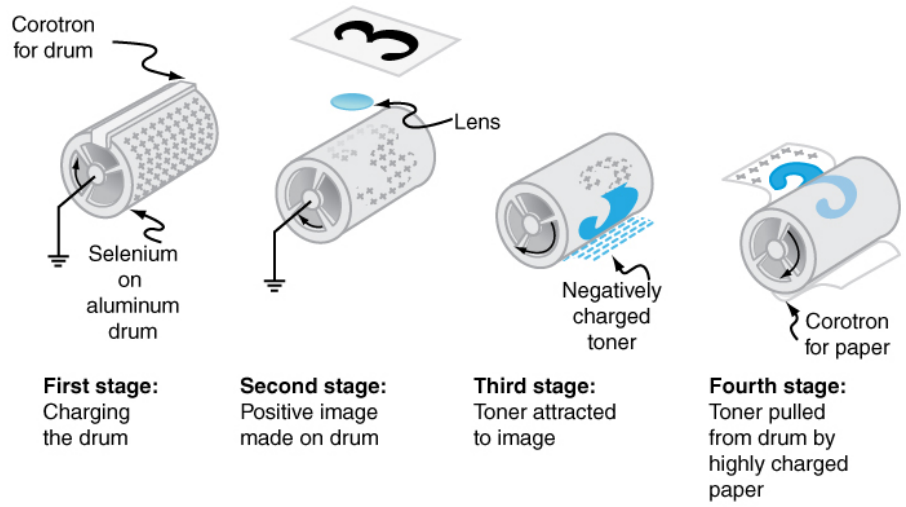
Xerography

Most copy machines use an electrostatic process called **xerography**—a word coined from the Greek words *xeros* for dry and *graphos* for writing. The heart of the process is shown in simplified form in [link].

A selenium-coated aluminum drum is sprayed with positive charge from points on a device called a corotron. Selenium is a substance with an interesting property—it is a **photoconductor**. That is, selenium is an insulator when in the dark and a conductor when exposed to light.

In the first stage of the xerography process, the conducting aluminum drum is **grounded** so that a negative charge is induced under the thin layer of uniformly positively charged selenium. In the second stage, the surface of the drum is exposed to the image of whatever is to be copied. Where the image is light, the selenium becomes conducting, and the positive charge is neutralized. In dark areas, the positive charge remains, and so the image has been transferred to the drum.

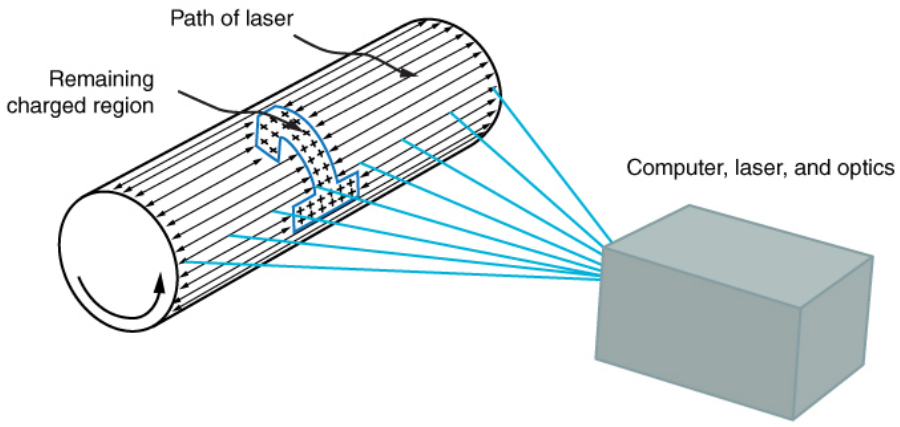
The third stage takes a dry black powder, called toner, and sprays it with a negative charge so that it will be attracted to the positive regions of the drum. Next, a blank piece of paper is given a greater positive charge than on the drum so that it will pull the toner from the drum. Finally, the paper and electrostatically held toner are passed through heated pressure rollers, which melt and permanently adhere the toner within the fibers of the paper.



Xerography is a dry copying process based on electrostatics. The major steps in the process are the charging of the photoconducting drum, transfer of an image creating a positive charge duplicate, attraction of toner to the charged parts of the drum, and transfer of toner to the paper. Not shown are heat treatment of the paper and cleansing of the drum for the next copy.

Laser Printers

Laser printers use the xerographic process to make high-quality images on paper, employing a laser to produce an image on the photoconducting drum as shown in [link]. In its most common application, the laser printer receives output from a computer, and it can achieve high-quality output because of the precision with which laser light can be controlled. Many laser printers do significant information processing, such as making sophisticated letters or fonts, and may contain a computer more powerful than the one giving them the raw data to be printed.

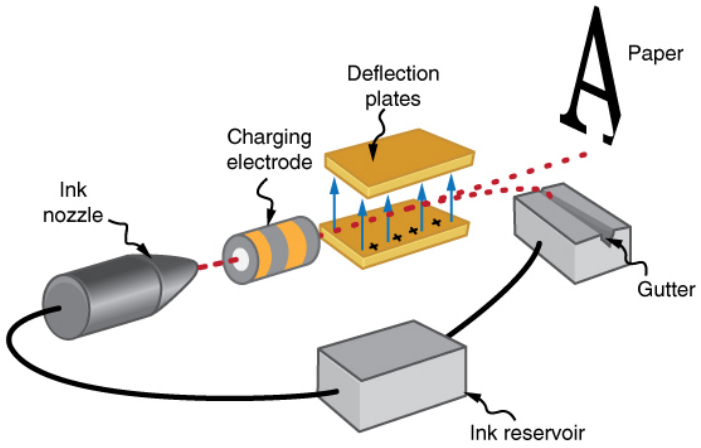


In a laser printer, a laser beam is scanned across a photoconducting drum, leaving a positive charge image. The other steps for charging the drum and transferring the image to paper are the same as in xerography. Laser light can be very precisely controlled, enabling laser printers to produce high-quality images.

Ink Jet Printers and Electrostatic Painting

The **ink jet printer**, commonly used to print computer-generated text and graphics, also employs electrostatics. A nozzle makes a fine spray of tiny ink droplets, which are then given an electrostatic charge. (See [link].)

Once charged, the droplets can be directed, using pairs of charged plates, with great precision to form letters and images on paper. Ink jet printers can produce color images by using a black jet and three other jets with primary colors, usually cyan, magenta, and yellow, much as a color television produces color. (This is more difficult with xerography, requiring multiple drums and toners.)



The nozzle of an ink-jet printer produces small ink droplets, which are sprayed with electrostatic charge. Various computer-driven devices are then used to direct the droplets to the correct positions on a page.

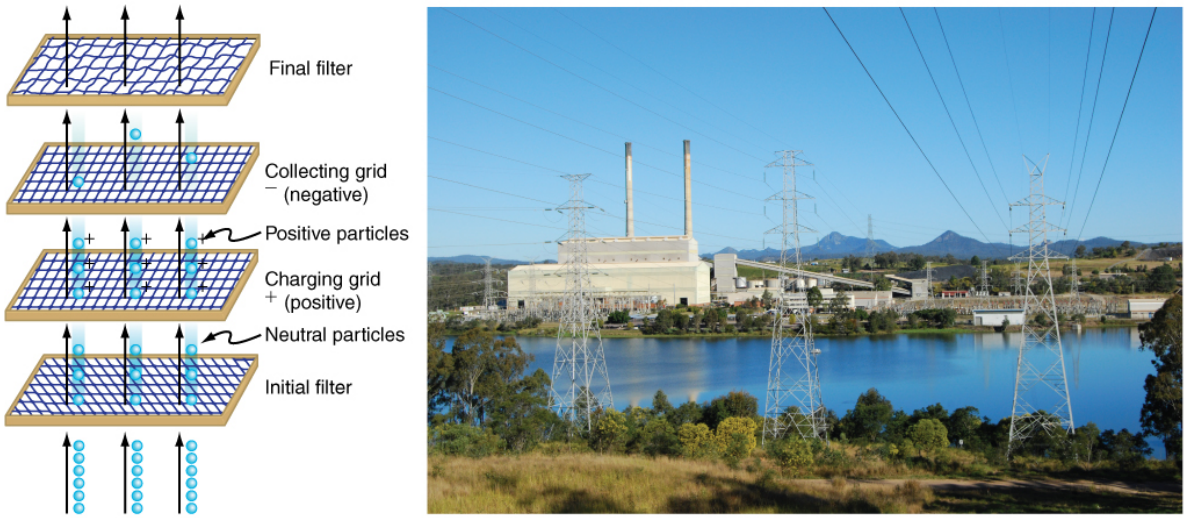
Electrostatic painting employs electrostatic charge to spray paint onto odd-shaped surfaces. Mutual repulsion of like charges causes the paint to fly away from its source. Surface tension forms drops, which are then attracted by unlike charges to the surface to be painted. Electrostatic painting can reach those hard-to-get at places, applying an even coat in a controlled

manner. If the object is a conductor, the electric field is perpendicular to the surface, tending to bring the drops in perpendicularly. Corners and points on conductors will receive extra paint. Felt can similarly be applied.

Smoke Precipitators and Electrostatic Air Cleaning

Another important application of electrostatics is found in air cleaners, both large and small. The electrostatic part of the process places excess (usually positive) charge on smoke, dust, pollen, and other particles in the air and then passes the air through an oppositely charged grid that attracts and retains the charged particles. (See [link].)

Large **electrostatic precipitators** are used industrially to remove over 99% of the particles from stack gas emissions associated with the burning of coal and oil. Home precipitators, often in conjunction with the home heating and air conditioning system, are very effective in removing polluting particles, irritants, and allergens.



(a) Schematic of an electrostatic precipitator. Air is passed through grids of opposite charge. The first grid charges airborne particles, while the second attracts and collects them. (b) The dramatic effect of

electrostatic precipitators is seen by the absence of smoke from this power plant. (credit: Cmdalgleish, Wikimedia Commons)

Note:
Problem-Solving Strategies for Electrostatics

1. Examine the situation to determine if static electricity is involved. This may concern separated stationary charges, the forces among them, and the electric fields they create.
2. Identify the system of interest. This includes noting the number, locations, and types of charges involved.
3. Identify exactly what needs to be determined in the problem (identify the unknowns). A written list is useful. Determine whether the Coulomb force is to be considered directly—if so, it may be useful to draw a free-body diagram, using electric field lines.
4. Make a list of what is given or can be inferred from the problem as stated (identify the knowns). It is important to distinguish the Coulomb force F from the electric field E, for example.
5. Solve the appropriate equation for the quantity to be determined (the unknown) or draw the field lines as requested.
6. Examine the answer to see if it is reasonable: Does it make sense? Are units correct and the numbers involved reasonable?

Integrated Concepts

The Integrated Concepts exercises for this module involve concepts such as electric charges, electric fields, and several other topics. Physics is most interesting when applied to general situations involving more than a narrow set of physical principles. The electric field exerts force on charges, for example, and hence the relevance of Dynamics: Force and Newton's Laws of Motion. The following topics are involved in some or all of the problems labeled "Integrated Concepts":

- Kinematics
- Two-Dimensional Kinematics
- Dynamics: Force and Newton’s Laws of Motion
- Uniform Circular Motion and Gravitation
- Statics and Torque
- Fluid Statics

The following worked example illustrates how this strategy is applied to an Integrated Concept problem:

Example:
Acceleration of a Charged Drop of Gasoline
If steps are not taken to ground a gasoline pump, static electricity can be placed on gasoline when filling your car’s tank. Suppose a tiny drop of gasoline has a mass of 4.00×10^{-15} kg and is given a positive charge of 3.20×10^{-19} C. (a) Find the weight of the drop. (b) Calculate the electric force on the drop if there is an upward electric field of strength 3.00×10^{5} N/C due to other static electricity in the vicinity. (c) Calculate the drop’s acceleration.
Strategy
To solve an integrated concept problem, we must first identify the physical principles involved and identify the chapters in which they are found. Part (a) of this example asks for weight. This is a topic of dynamics and is defined in Dynamics: Force and Newton’s Laws of Motion. Part (b) deals with electric force on a charge, a topic of Electric Charge and Electric Field. Part (c) asks for acceleration, knowing forces and mass. These are part of Newton’s laws, also found in Dynamics: Force and Newton’s Laws of Motion.
The following solutions to each part of the example illustrate how the specific problem-solving strategies are applied. These involve identifying knowns and unknowns, checking to see if the answer is reasonable, and so on.
Solution for (a)
Weight is mass times the acceleration due to gravity, as first expressed in **Equation:**

$$w = \mathrm{mg}.$$

Entering the given mass and the average acceleration due to gravity yields
Equation:

$$w = (4.00 \times 10^{-15}\ \mathrm{kg})(9.80\ \mathrm{m/s^2}) = 3.92 \times 10^{-14}\ \mathrm{N}.$$

Discussion for (a)
This is a small weight, consistent with the small mass of the drop.
Solution for (b)
The force an electric field exerts on a charge is given by rearranging the following equation:
Equation:

$$F = \mathrm{qE}.$$

Here we are given the charge ($3.20 \times 10^{-19}\ \mathrm{C}$ is twice the fundamental unit of charge) and the electric field strength, and so the electric force is found to be
Equation:

$$F = (3.20 \times 10^{-19}\ \mathrm{C})(3.00 \times 10^{5}\ \mathrm{N/C}) = 9.60 \times 10^{-14}\ \mathrm{N}.$$

Discussion for (b)
While this is a small force, it is greater than the weight of the drop.
Solution for (c)
The acceleration can be found using Newton’s second law, provided we can identify all of the external forces acting on the drop. We assume only the drop’s weight and the electric force are significant. Since the drop has a positive charge and the electric field is given to be upward, the electric force is upward. We thus have a one-dimensional (vertical direction) problem, and we can state Newton’s second law as
Equation:

$$a = \frac{F_{\mathrm{net}}}{m}.$$

where $F_{\text{net}} = F - w$. Entering this and the known values into the expression for Newton's second law yields

Equation:

$$\begin{aligned} a &= \frac{F-w}{m} \\ &= \frac{9.60\times10^{-14}\ \text{N}-3.92\times10^{-14}\ \text{N}}{4.00\times10^{-15}\ \text{kg}} \\ &= 14.2\ \text{m/s}^2. \end{aligned}$$

Discussion for (c)

This is an upward acceleration great enough to carry the drop to places where you might not wish to have gasoline.

This worked example illustrates how to apply problem-solving strategies to situations that include topics in different chapters. The first step is to identify the physical principles involved in the problem. The second step is to solve for the unknown using familiar problem-solving strategies. These are found throughout the text, and many worked examples show how to use them for single topics. In this integrated concepts example, you can see how to apply them across several topics. You will find these techniques useful in applications of physics outside a physics course, such as in your profession, in other science disciplines, and in everyday life. The following problems will build your skills in the broad application of physical principles.

Note:

Unreasonable Results

The Unreasonable Results exercises for this module have results that are unreasonable because some premise is unreasonable or because certain of the premises are inconsistent with one another. Physical principles applied correctly then produce unreasonable results. The purpose of these problems is to give practice in assessing whether nature is being accurately described, and if it is not to trace the source of difficulty.

Note:
Problem-Solving Strategy
To determine if an answer is reasonable, and to determine the cause if it is not, do the following.

1. Solve the problem using strategies as outlined above. Use the format followed in the worked examples in the text to solve the problem as usual.
2. Check to see if the answer is reasonable. Is it too large or too small, or does it have the wrong sign, improper units, and so on?
3. If the answer is unreasonable, look for what specifically could cause the identified difficulty. Usually, the manner in which the answer is unreasonable is an indication of the difficulty. For example, an extremely large Coulomb force could be due to the assumption of an excessively large separated charge.

Section Summary

- Electrostatics is the study of electric fields in static equilibrium.
- In addition to research using equipment such as a Van de Graaff generator, many practical applications of electrostatics exist, including photocopiers, laser printers, ink-jet printers and electrostatic air filters.

Problems & Exercises

Exercise:

Problem:

(a) What is the electric field 5.00 m from the center of the terminal of a Van de Graaff with a 3.00 mC charge, noting that the field is equivalent to that of a point charge at the center of the terminal? (b) At this distance, what force does the field exert on a $2.00\ \mu\text{C}$ charge on the Van de Graaff's belt?

Exercise:

Problem:

(a) What is the direction and magnitude of an electric field that supports the weight of a free electron near the surface of Earth? (b) Discuss what the small value for this field implies regarding the relative strength of the gravitational and electrostatic forces.

Solution:

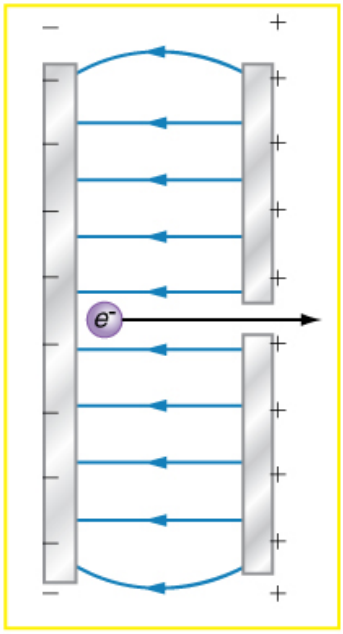
(a) $5.58 \times 10^{-11}\ \text{N/C}$

(b)the coulomb force is extraordinarily stronger than gravity

Exercise:

Problem:

A simple and common technique for accelerating electrons is shown in [link], where there is a uniform electric field between two plates. Electrons are released, usually from a hot filament, near the negative plate, and there is a small hole in the positive plate that allows the electrons to continue moving. (a) Calculate the acceleration of the electron if the field strength is $2.50 \times 10^{4}\ \text{N/C}$. (b) Explain why the electron will not be pulled back to the positive plate once it moves through the hole.



Parallel conducting plates with opposite charges on them create a relatively uniform electric field used to accelerate electrons to the right. Those that go through the hole can be used to make a TV or computer screen glow or to produce X-rays.

Exercise:

Problem:

Earth has a net charge that produces an electric field of approximately 150 N/C downward at its surface. (a) What is the magnitude and sign of the excess charge, noting the electric field of a conducting sphere is equivalent to a point charge at its center? (b) What acceleration will the field produce on a free electron near Earth's surface? (c) What mass object with a single extra electron will have its weight supported by this field?

Solution:

(a) $-6.76 \times 10^5\ \text{C}$

(b) $2.63 \times 10^{13}\ \text{m/s}^2\ (\text{upward})$

(c) $2.45 \times 10^{-18}\ \text{kg}$

Exercise:

Problem:

Point charges of $25.0\ \mu\text{C}$ and $45.0\ \mu\text{C}$ are placed 0.500 m apart. (a) At what point along the line between them is the electric field zero? (b) What is the electric field halfway between them?

Exercise:

Problem:

What can you say about two charges q_1 and q_2, if the electric field one-fourth of the way from q_1 to q_2 is zero?

Solution:

The charge q_2 is 9 times greater than q_1.

Exercise:

Problem: **Integrated Concepts**

Calculate the angular velocity ω of an electron orbiting a proton in the hydrogen atom, given the radius of the orbit is 0.530×10^{-10} m. You may assume that the proton is stationary and the centripetal force is supplied by Coulomb attraction.

Exercise:

Problem: **Integrated Concepts**

An electron has an initial velocity of 5.00×10^{6} m/s in a uniform 2.00×10^{5} N/C strength electric field. The field accelerates the electron in the direction opposite to its initial velocity. (a) What is the direction of the electric field? (b) How far does the electron travel before coming to rest? (c) How long does it take the electron to come to rest? (d) What is the electron's velocity when it returns to its starting point?

Exercise:

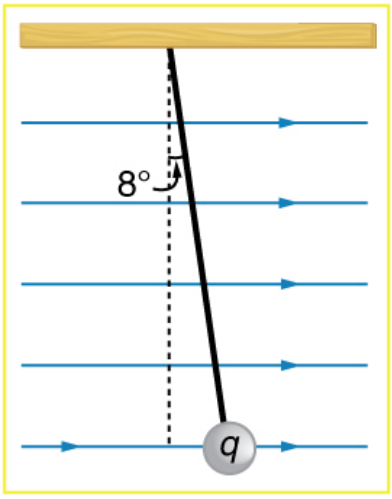
Problem: **Integrated Concepts**

The practical limit to an electric field in air is about 3.00×10^{6} N/C. Above this strength, sparking takes place because air begins to ionize and charges flow, reducing the field. (a) Calculate the distance a free proton must travel in this field to reach 3.00% of the speed of light, starting from rest. (b) Is this practical in air, or must it occur in a vacuum?

Exercise:

Problem: **Integrated Concepts**

A 5.00 g charged insulating ball hangs on a 30.0 cm long string in a uniform horizontal electric field as shown in [link]. Given the charge on the ball is 1.00 μC, find the strength of the field.

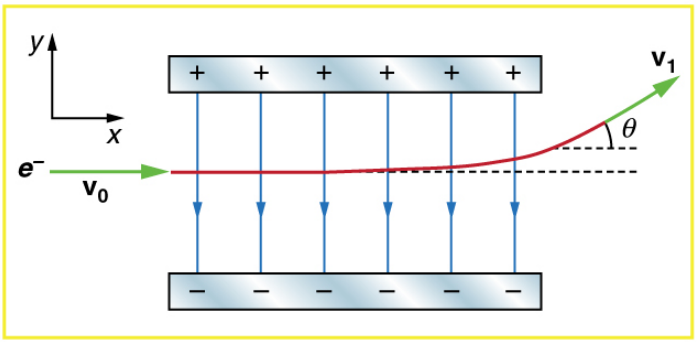


A horizontal electric field causes the charged ball to hang at an angle of 8.00º.

Exercise:

Problem: Integrated Concepts

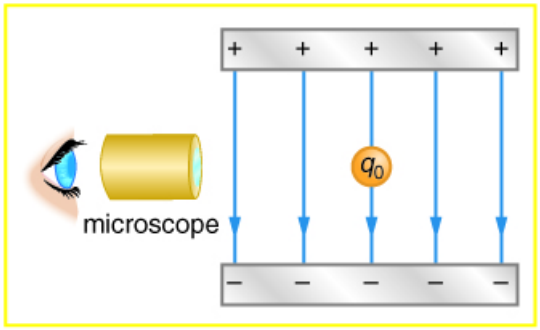
[link] shows an electron passing between two charged metal plates that create an 100 N/C vertical electric field perpendicular to the electron's original horizontal velocity. (These can be used to change the electron's direction, such as in an oscilloscope.) The initial speed of the electron is 3.00×10^6 m/s, and the horizontal distance it travels in the uniform field is 4.00 cm. (a) What is its vertical deflection? (b) What is the vertical component of its final velocity? (c) At what angle does it exit? Neglect any edge effects.



Exercise:

Problem: **Integrated Concepts**

The classic Millikan oil drop experiment was the first to obtain an accurate measurement of the charge on an electron. In it, oil drops were suspended against the gravitational force by a vertical electric field. (See [link].) Given the oil drop to be 1.00 μm in radius and have a density of 920 kg/m^3: (a) Find the weight of the drop. (b) If the drop has a single excess electron, find the electric field strength needed to balance its weight.



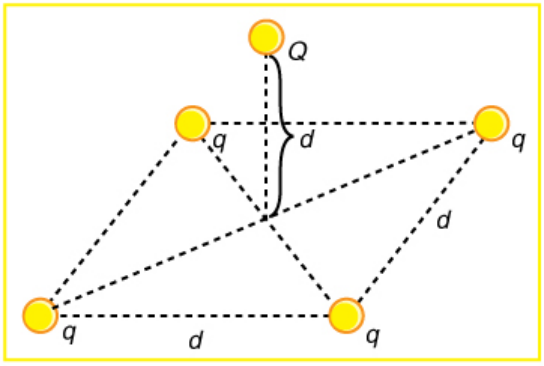
In the Millikan oil drop experiment, small drops can be suspended in an electric field by the force exerted on a single excess electron. Classically, this experiment was used to determine the electron charge q_e by

measuring the electric field and mass of the drop.

Exercise:

Problem: Integrated Concepts

(a) In [link], four equal charges q lie on the corners of a square. A fifth charge Q is on a mass m directly above the center of the square, at a height equal to the length d of one side of the square. Determine the magnitude of q in terms of Q, m, and d, if the Coulomb force is to equal the weight of m. (b) Is this equilibrium stable or unstable? Discuss.



Four equal charges on the corners of a horizontal square support the weight of a fifth charge located directly above the center of the square.

Exercise:

Problem: Unreasonable Results

(a) Calculate the electric field strength near a 10.0 cm diameter conducting sphere that has 1.00 C of excess charge on it. (b) What is

unreasonable about this result? (c) Which assumptions are responsible?

Exercise:

Problem: Unreasonable Results

(a) Two 0.500 g raindrops in a thunderhead are 1.00 cm apart when they each acquire 1.00 mC charges. Find their acceleration. (b) What is unreasonable about this result? (c) Which premise or assumption is responsible?

Exercise:

Problem: Unreasonable Results

A wrecking yard inventor wants to pick up cars by charging a 0.400 m diameter ball and inducing an equal and opposite charge on the car. If a car has a 1000 kg mass and the ball is to be able to lift it from a distance of 1.00 m: (a) What minimum charge must be used? (b) What is the electric field near the surface of the ball? (c) Why are these results unreasonable? (d) Which premise or assumption is responsible?

Exercise:

Problem: Construct Your Own Problem

Consider two insulating balls with evenly distributed equal and opposite charges on their surfaces, held with a certain distance between the centers of the balls. Construct a problem in which you calculate the electric field (magnitude and direction) due to the balls at various points along a line running through the centers of the balls and extending to infinity on either side. Choose interesting points and comment on the meaning of the field at those points. For example, at what points might the field be just that due to one ball and where does the field become negligibly small? Among the things to be considered are the magnitudes of the charges and the distance between the centers of the balls. Your instructor may wish for you to consider the electric

field off axis or for a more complex array of charges, such as those in a water molecule.

Exercise:

Problem: Construct Your Own Problem

Consider identical spherical conducting space ships in deep space where gravitational fields from other bodies are negligible compared to the gravitational attraction between the ships. Construct a problem in which you place identical excess charges on the space ships to exactly counter their gravitational attraction. Calculate the amount of excess charge needed. Examine whether that charge depends on the distance between the centers of the ships, the masses of the ships, or any other factors. Discuss whether this would be an easy, difficult, or even impossible thing to do in practice.

Glossary

Van de Graaff generator
: a machine that produces a large amount of excess charge, used for experiments with high voltage

electrostatics
: the study of electric forces that are static or slow-moving

photoconductor
: a substance that is an insulator until it is exposed to light, when it becomes a conductor

xerography
: a dry copying process based on electrostatics

grounded
: connected to the ground with a conductor, so that charge flows freely to and from the Earth to the grounded object

laser printer
: uses a laser to create a photoconductive image on a drum, which attracts dry ink particles that are then rolled onto a sheet of paper to print a high-quality copy of the image

ink-jet printer
: small ink droplets sprayed with an electric charge are controlled by electrostatic plates to create images on paper

electrostatic precipitators
: filters that apply charges to particles in the air, then attract those charges to a filter, removing them from the airstream

Introduction to Electric Current, Resistance, and Ohm's Law

class="introduction"

Electric energy in massive quantities is transmitted from this hydroelectric facility, the Srisailam power station located along the Krishna River in India, by the movement of charge—that is, by electric current. (credit: Chintohere, Wikimedia Commons)



The flicker of numbers on a handheld calculator, nerve impulses carrying signals of vision to the brain, an ultrasound device sending a signal to a computer screen, the brain sending a message for a baby to twitch its toes, an electric train pulling its load over a mountain pass, a hydroelectric plant sending energy to metropolitan and rural users—these and many other examples of electricity involve *electric current, the movement of charge*. Humankind has indeed harnessed electricity, the basis of technology, to improve our quality of life. Whereas the previous two chapters concentrated on static electricity and the fundamental force underlying its behavior, the next few chapters will be devoted to electric and magnetic phenomena involving current. In addition to exploring applications of electricity, we shall gain new insights into nature—in particular, the fact that all magnetism results from electric current.

Introduction to Electric Potential and Electric Energy
class="introduction"

Automated external defibrillator unit (AED) (credit: U.S. Defense Department photo/Tech. Sgt. Suzanne M. Day)



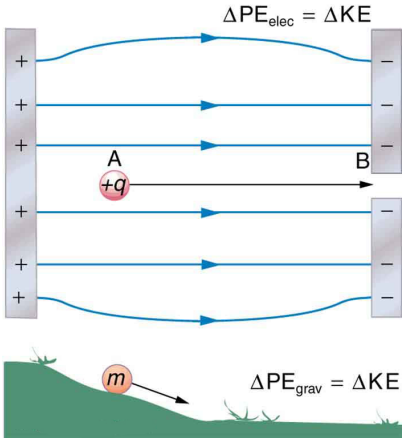
In Electric Charge and Electric Field, we just scratched the surface (or at least rubbed it) of electrical phenomena. Two of the most familiar aspects of

electricity are its energy and *voltage*. We know, for example, that great amounts of electrical energy can be stored in batteries, are transmitted cross-country through power lines, and may jump from clouds to explode the sap of trees. In a similar manner, at molecular levels, *ions* cross cell membranes and transfer information. We also know about voltages associated with electricity. Batteries are typically a few volts, the outlets in your home produce 120 volts, and power lines can be as high as hundreds of thousands of volts. But energy and voltage are not the same thing. A motorcycle battery, for example, is small and would not be very successful in replacing the much larger car battery, yet each has the same voltage. In this chapter, we shall examine the relationship between voltage and electrical energy and begin to explore some of the many applications of electricity.

Electric Potential Energy: Potential Difference

- Define electric potential and electric potential energy.
- Describe the relationship between potential difference and electrical potential energy.
- Explain electron volt and its usage in submicroscopic process.
- Determine electric potential energy given potential difference and amount of charge.

When a free positive charge q is accelerated by an electric field, such as shown in [link], it is given kinetic energy. The process is analogous to an object being accelerated by a gravitational field. It is as if the charge is going down an electrical hill where its electric potential energy is converted to kinetic energy. Let us explore the work done on a charge q by the electric field in this process, so that we may develop a definition of electric potential energy.



A charge accelerated by an electric field is analogous to a mass going down a hill. In both cases potential energy is converted to another form. Work is done by a force, but since this force

is conservative, we can write $W = -\Delta\text{PE}$.

The electrostatic or Coulomb force is conservative, which means that the work done on *q* is independent of the path taken. This is exactly analogous to the gravitational force in the absence of dissipative forces such as friction. When a force is conservative, it is possible to define a potential energy associated with the force, and it is usually easier to deal with the potential energy (because it depends only on position) than to calculate the work directly.

We use the letters PE to denote electric potential energy, which has units of joules (J). The change in potential energy, ΔPE, is crucial, since the work done by a conservative force is the negative of the change in potential energy; that is, $W = -\Delta\text{PE}$. For example, work W done to accelerate a positive charge from rest is positive and results from a loss in PE, or a negative ΔPE. There must be a minus sign in front of ΔPE to make W positive. PE can be found at any point by taking one point as a reference and calculating the work needed to move a charge to the other point.

Note:
Potential Energy
$W = -\Delta\text{PE}$. For example, work W done to accelerate a positive charge from rest is positive and results from a loss in PE, or a negative ΔPE. There must be a minus sign in front of ΔPE to make W positive. PE can be found at any point by taking one point as a reference and calculating the work needed to move a charge to the other point.

Gravitational potential energy and electric potential energy are quite analogous. Potential energy accounts for work done by a conservative force and gives added insight regarding energy and energy transformation

without the necessity of dealing with the force directly. It is much more common, for example, to use the concept of voltage (related to electric potential energy) than to deal with the Coulomb force directly.

Calculating the work directly is generally difficult, since $W = \mathrm{Fd}\cos\theta$ and the direction and magnitude of F can be complex for multiple charges, for odd-shaped objects, and along arbitrary paths. But we do know that, since $F = \mathrm{qE}$, the work, and hence $\Delta\mathrm{PE}$, is proportional to the test charge q. To have a physical quantity that is independent of test charge, we define **electric potential** V (or simply potential, since electric is understood) to be the potential energy per unit charge:
Equation:

$$V = \frac{\mathrm{PE}}{q}.$$

Note:
Electric Potential
This is the electric potential energy per unit charge.
Equation:

$$V = \frac{\mathrm{PE}}{q}$$

Since PE is proportional to *q* , the dependence on *q* cancels. Thus V does not depend on *q*. The change in potential energy $\Delta\mathrm{PE}$ is crucial, and so we are concerned with the difference in potential or potential difference ΔV between two points, where
Equation:

$$\Delta V = V_{\mathrm{B}} - V_{\mathrm{A}} = \frac{\Delta\mathrm{PE}}{q}.$$

The **potential difference** between points A and B, $V_\mathrm{B} - V_\mathrm{A}$, is thus defined to be the change in potential energy of a charge q moved from A to B, divided by the charge. Units of potential difference are joules per coulomb, given the name volt (V) after Alessandro Volta.
Equation:

$$1\ \mathrm{V} = 1\ \frac{\mathrm{J}}{\mathrm{C}}$$

Note:
Potential Difference
The potential difference between points A and B, $V_\mathrm{B} - V_\mathrm{A}$, is defined to be the change in potential energy of a charge q moved from A to B, divided by the charge. Units of potential difference are joules per coulomb, given the name volt (V) after Alessandro Volta.
Equation:

$$1\ \mathrm{V} = 1\ \frac{\mathrm{J}}{\mathrm{C}}$$

The familiar term **voltage** is the common name for potential difference. Keep in mind that whenever a voltage is quoted, it is understood to be the potential difference between two points. For example, every battery has two terminals, and its voltage is the potential difference between them. More fundamentally, the point you choose to be zero volts is arbitrary. This is analogous to the fact that gravitational potential energy has an arbitrary zero, such as sea level or perhaps a lecture hall floor.

In summary, the relationship between potential difference (or voltage) and electrical potential energy is given by
Equation:

$$\Delta V = \frac{\Delta \text{PE}}{q} \text{ and } \Delta \text{PE} = q\Delta V.$$

Note:
Potential Difference and Electrical Potential Energy
The relationship between potential difference (or voltage) and electrical potential energy is given by
Equation:

$$\Delta V = \frac{\Delta \text{PE}}{q} \text{ and } \Delta \text{PE} = q\Delta V.$$

The second equation is equivalent to the first.

Voltage is not the same as energy. Voltage is the energy per unit charge. Thus a motorcycle battery and a car battery can both have the same voltage (more precisely, the same potential difference between battery terminals), yet one stores much more energy than the other since $\Delta \text{PE} = q\Delta V$. The car battery can move more charge than the motorcycle battery, although both are 12 V batteries.

Example:
Calculating Energy
Suppose you have a 12.0 V motorcycle battery that can move 5000 C of charge, and a 12.0 V car battery that can move 60,000 C of charge. How much energy does each deliver? (Assume that the numerical value of each charge is accurate to three significant figures.)
Strategy
To say we have a 12.0 V battery means that its terminals have a 12.0 V potential difference. When such a battery moves charge, it puts the charge

through a potential difference of 12.0 V, and the charge is given a change in potential energy equal to $\Delta \text{PE} = q\Delta V$.
So to find the energy output, we multiply the charge moved by the potential difference.

Solution

For the motorcycle battery, $q = 5000\ \text{C}$ and $\Delta V = 12.0\ \text{V}$. The total energy delivered by the motorcycle battery is

Equation:

$$\begin{aligned} \Delta \text{PE}_{\text{cycle}} &= (5000\ \text{C})(12.0\ \text{V}) \\ &= (5000\ \text{C})(12.0\ \text{J/C}) \\ &= 6.00 \times 10^4\ \text{J}. \end{aligned}$$

Similarly, for the car battery, $q = 60{,}000\ \text{C}$ and

Equation:

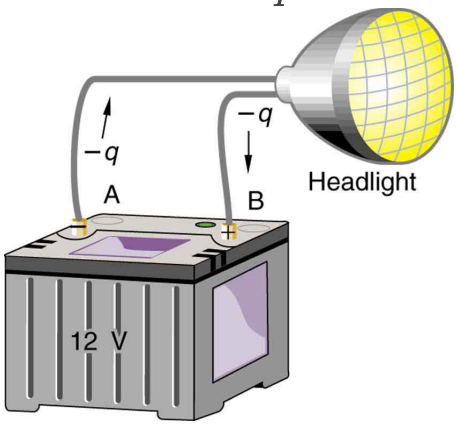
$$\begin{aligned} \Delta \text{PE}_{\text{car}} &= (60{,}000\ \text{C})(12.0\ \text{V}) \\ &= 7.20 \times 10^5\ \text{J}. \end{aligned}$$

Discussion

While voltage and energy are related, they are not the same thing. The voltages of the batteries are identical, but the energy supplied by each is quite different. Note also that as a battery is discharged, some of its energy is used internally and its terminal voltage drops, such as when headlights dim because of a low car battery. The energy supplied by the battery is still calculated as in this example, but not all of the energy is available for external use.

Note that the energies calculated in the previous example are absolute values. The change in potential energy for the battery is negative, since it loses energy. These batteries, like many electrical systems, actually move negative charge—electrons in particular. The batteries repel electrons from their negative terminals (A) through whatever circuitry is involved and attract them to their positive terminals (B) as shown in [link]. The change in potential is $\Delta V = V_{\text{B}} - V_{\text{A}} = +12\ \text{V}$ and the charge q is negative, so that

$\Delta \text{PE} = q\Delta V$ is negative, meaning the potential energy of the battery has decreased when q has moved from A to B.



A battery moves negative charge from its negative terminal through a headlight to its positive terminal. Appropriate combinations of chemicals in the battery separate charges so that the negative terminal has an excess of negative charge, which is repelled by it and attracted to the excess positive charge on the other terminal. In terms of potential, the positive terminal is at a higher voltage than the negative. Inside the battery, both positive and negative charges move.

Example:

How Many Electrons Move through a Headlight Each Second?

When a 12.0 V car battery runs a single 30.0 W headlight, how many electrons pass through it each second?

Strategy

To find the number of electrons, we must first find the charge that moved in 1.00 s. The charge moved is related to voltage and energy through the equation $\Delta\text{PE} = q\Delta V$. A 30.0 W lamp uses 30.0 joules per second. Since the battery loses energy, we have $\Delta\text{PE} = –30.0\ \text{J}$ and, since the electrons are going from the negative terminal to the positive, we see that $\Delta V = +12.0\ \text{V}$.

Solution

To find the charge q moved, we solve the equation $\Delta\text{PE} = q\Delta V$:

Equation:

$$q = \frac{\Delta\text{PE}}{\Delta V}.$$

Entering the values for ΔPE and ΔV, we get

Equation:

$$q = \frac{–30.0\ \text{J}}{+12.0\ \text{V}} = \frac{–30.0\ \text{J}}{+12.0\ \text{J/C}} = –2.50\ \text{C}.$$

The number of electrons n_e is the total charge divided by the charge per electron. That is,

Equation:

$$\text{n}_\text{e} = \frac{–2.50\ \text{C}}{–1.60\times10^{-19}\ \text{C/e}^-} = 1.56\times10^{19}\ \text{electrons}.$$

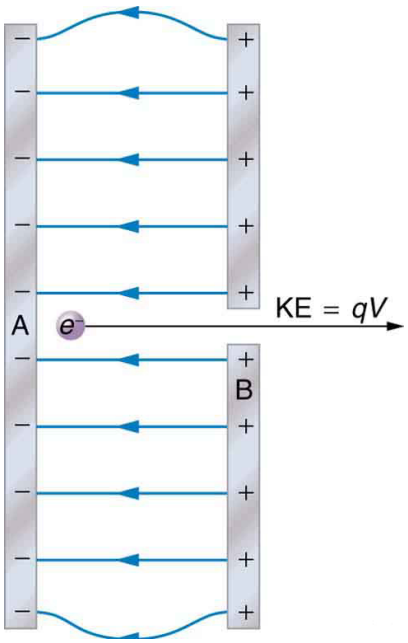
Discussion

This is a very large number. It is no wonder that we do not ordinarily observe individual electrons with so many being present in ordinary

systems. In fact, electricity had been in use for many decades before it was determined that the moving charges in many circumstances were negative. Positive charge moving in the opposite direction of negative charge often produces identical effects; this makes it difficult to determine which is moving or whether both are moving.

The Electron Volt

The energy per electron is very small in macroscopic situations like that in the previous example—a tiny fraction of a joule. But on a submicroscopic scale, such energy per particle (electron, proton, or ion) can be of great importance. For example, even a tiny fraction of a joule can be great enough for these particles to destroy organic molecules and harm living tissue. The particle may do its damage by direct collision, or it may create harmful x rays, which can also inflict damage. It is useful to have an energy unit related to submicroscopic effects. [link] shows a situation related to the definition of such an energy unit. An electron is accelerated between two charged metal plates as it might be in an old-model television tube or oscilloscope. The electron is given kinetic energy that is later converted to another form—light in the television tube, for example. (Note that downhill for the electron is uphill for a positive charge.) Since energy is related to voltage by $\Delta \text{PE} = q\Delta V$, we can think of the joule as a coulomb-volt.



A typical electron gun accelerates electrons using a potential difference between two metal plates. The energy of the electron in electron volts is numerically the same as the voltage between the plates. For example, a 5000 V potential difference produces 5000 eV electrons.

On the submicroscopic scale, it is more convenient to define an energy unit called the **electron volt** (eV), which is the energy given to a fundamental charge accelerated through a potential difference of 1 V. In equation form,
Equation:

$$\begin{aligned} 1\text{ eV} &= \left(1.60 \times 10^{-19}\text{ C}\right)(1\text{ V}) = \left(1.60 \times 10^{-19}\text{ C}\right)(1\text{ J/C}) \\ &= 1.60 \times 10^{-19}\text{ J}. \end{aligned}$$

Note:
Electron Volt
On the submicroscopic scale, it is more convenient to define an energy unit called the electron volt (eV), which is the energy given to a fundamental charge accelerated through a potential difference of 1 V. In equation form,
Equation:

$$\begin{aligned} 1\text{ eV} &= \left(1.60 \times 10^{-19}\text{ C}\right)(1\text{ V}) = \left(1.60 \times 10^{-19}\text{ C}\right)(1\text{ J/C}) \\ &= 1.60 \times 10^{-19}\text{ J}. \end{aligned}$$

An electron accelerated through a potential difference of 1 V is given an energy of 1 eV. It follows that an electron accelerated through 50 V is given 50 eV. A potential difference of 100,000 V (100 kV) will give an electron an energy of 100,000 eV (100 keV), and so on. Similarly, an ion with a double positive charge accelerated through 100 V will be given 200 eV of energy. These simple relationships between accelerating voltage and particle charges make the electron volt a simple and convenient energy unit in such circumstances.

Note:
Connections: Energy Units
The electron volt (eV) is the most common energy unit for submicroscopic processes. This will be particularly noticeable in the chapters on modern physics. Energy is so important to so many subjects that there is a tendency to define a special energy unit for each major topic. There are, for example,

calories for food energy, kilowatt-hours for electrical energy, and therms for natural gas energy.

The electron volt is commonly employed in submicroscopic processes—chemical valence energies and molecular and nuclear binding energies are among the quantities often expressed in electron volts. For example, about 5 eV of energy is required to break up certain organic molecules. If a proton is accelerated from rest through a potential difference of 30 kV, it is given an energy of 30 keV (30,000 eV) and it can break up as many as 6000 of these molecules ($30{,}000\ \text{eV} \div 5\ \text{eV per molecule} = 6000\ \text{molecules}$). Nuclear decay energies are on the order of 1 MeV (1,000,000 eV) per event and can, thus, produce significant biological damage.

Conservation of Energy

The total energy of a system is conserved if there is no net addition (or subtraction) of work or heat transfer. For conservative forces, such as the electrostatic force, conservation of energy states that mechanical energy is a constant.

Mechanical energy is the sum of the kinetic energy and potential energy of a system; that is, $\text{KE} + \text{PE} = \text{constant}$. A loss of PE of a charged particle becomes an increase in its KE. Here PE is the electric potential energy. Conservation of energy is stated in equation form as
Equation:

$$\text{KE} + \text{PE} = \text{constant}$$

or
Equation:

$$\text{KE}_\text{i} + \text{PE}_\text{i} = \text{KE}_\text{f} + \text{PE}_\text{f},$$

where i and f stand for initial and final conditions. As we have found many times before, considering energy can give us insights and facilitate problem solving.

Example:
Electrical Potential Energy Converted to Kinetic Energy
Calculate the final speed of a free electron accelerated from rest through a potential difference of 100 V. (Assume that this numerical value is accurate to three significant figures.)
Strategy
We have a system with only conservative forces. Assuming the electron is accelerated in a vacuum, and neglecting the gravitational force (we will check on this assumption later), all of the electrical potential energy is converted into kinetic energy. We can identify the initial and final forms of energy to be $\mathrm{KE_i} = 0$, $\mathrm{KE_f} = ½\,mv^2$, $\mathrm{PE_i} = qV$, and $\mathrm{PE_f} = 0$.
Solution
Conservation of energy states that
Equation:

$$\mathrm{KE_i} + \mathrm{PE_i} = \mathrm{KE_f} + \mathrm{PE_f}.$$

Entering the forms identified above, we obtain
Equation:

$$qV = \frac{mv^2}{2}.$$

We solve this for v*:*
Equation:

$$v = \sqrt{\frac{2\mathrm{qV}}{m}}.$$

Entering values for q, V, and m gives
Equation:

$$\begin{aligned} v &= \sqrt{\frac{2\left(-1.60\times10^{-19}\text{ C}\right)\left(-100\text{ J/C}\right)}{9.11\times10^{-31}\text{ kg}}} \\ &= 5.93\times10^{6}\text{ m/s}. \end{aligned}$$

Discussion

Note that both the charge and the initial voltage are negative, as in [link]. From the discussions in Electric Charge and Electric Field, we know that electrostatic forces on small particles are generally very large compared with the gravitational force. The large final speed confirms that the gravitational force is indeed negligible here. The large speed also indicates how easy it is to accelerate electrons with small voltages because of their very small mass. Voltages much higher than the 100 V in this problem are typically used in electron guns. Those higher voltages produce electron speeds so great that relativistic effects must be taken into account. That is why a low voltage is considered (accurately) in this example.

Section Summary

- Electric potential is potential energy per unit charge.
- The potential difference between points A and B, $V_{\text{B}}-V_{\text{A}}$, defined to be the change in potential energy of a charge q moved from A to B, is equal to the change in potential energy divided by the charge, Potential difference is commonly called voltage, represented by the symbol ΔV.
  Equation:

$$\Delta V = \frac{\Delta\text{PE}}{q} \text{ and } \Delta\text{PE} = q\Delta V.$$

- An electron volt is the energy given to a fundamental charge accelerated through a potential difference of 1 V. In equation form,
  Equation:

$$\begin{aligned} 1\text{ eV} &= \left(1.60\times10^{-19}\text{ C}\right)(1\text{ V}) = \left(1.60\times10^{-19}\text{ C}\right)(1\text{ J/C}) \\ &= 1.60\times10^{-19}\text{ J}. \end{aligned}$$

- Mechanical energy is the sum of the kinetic energy and potential energy of a system, that is, $\mathrm{KE} + \mathrm{PE}$. This sum is a constant.

Conceptual Questions

Exercise:

Problem:

Voltage is the common word for potential difference. Which term is more descriptive, voltage or potential difference?

Exercise:

Problem:

If the voltage between two points is zero, can a test charge be moved between them with zero net work being done? Can this necessarily be done without exerting a force? Explain.

Exercise:

Problem:

What is the relationship between voltage and energy? More precisely, what is the relationship between potential difference and electric potential energy?

Exercise:

Problem: Voltages are always measured between two points. Why?

Exercise:

Problem:

How are units of volts and electron volts related? How do they differ?

Problems & Exercises

Exercise:

Problem:

Find the ratio of speeds of an electron and a negative hydrogen ion (one having an extra electron) accelerated through the same voltage, assuming non-relativistic final speeds. Take the mass of the hydrogen ion to be 1.67×10^{-27} kg.

Solution:

42.8

Exercise:

Problem:

An evacuated tube uses an accelerating voltage of 40 kV to accelerate electrons to hit a copper plate and produce x rays. Non-relativistically, what would be the maximum speed of these electrons?

Exercise:

Problem:

A bare helium nucleus has two positive charges and a mass of 6.64×10^{-27} kg. (a) Calculate its kinetic energy in joules at 2.00% of the speed of light. (b) What is this in electron volts? (c) What voltage would be needed to obtain this energy?

Exercise:

Problem: Integrated Concepts

Singly charged gas ions are accelerated from rest through a voltage of 13.0 V. At what temperature will the average kinetic energy of gas molecules be the same as that given these ions?

Solution:

1.00×10^{5} K

Exercise:

Problem: Integrated Concepts

The temperature near the center of the Sun is thought to be 15 million degrees Celsius $\left(1.5\times10^{7}\ ^{\circ}\text{C}\right)$. Through what voltage must a singly charged ion be accelerated to have the same energy as the average kinetic energy of ions at this temperature?

Exercise:

Problem: Integrated Concepts

(a) What is the average power output of a heart defibrillator that dissipates 400 J of energy in 10.0 ms? (b) Considering the high-power output, why doesn’t the defibrillator produce serious burns?

Solution:

(a) $4\times10^{4}\ \text{W}$

(b) A defibrillator does not cause serious burns because the skin conducts electricity well at high voltages, like those used in defibrillators. The gel used aids in the transfer of energy to the body, and the skin doesn’t absorb the energy, but rather lets it pass through to the heart.

Exercise:

Problem: Integrated Concepts

A lightning bolt strikes a tree, moving 20.0 C of charge through a potential difference of $1.00\times10^{2}\ \text{MV}$. (a) What energy was dissipated? (b) What mass of water could be raised from 15°C to the boiling point and then boiled by this energy? (c) Discuss the damage that could be caused to the tree by the expansion of the boiling steam.

Exercise:

Problem: **Integrated Concepts**

A 12.0 V battery-operated bottle warmer heats 50.0 g of glass, 2.50×10^2 g of baby formula, and 2.00×10^2 g of aluminum from 20.0°C to 90.0°C. (a) How much charge is moved by the battery? (b) How many electrons per second flow if it takes 5.00 min to warm the formula? (Hint: Assume that the specific heat of baby formula is about the same as the specific heat of water.)

Solution:

(a) 7.40×10^3 C

(b) 1.54×10^{20} electrons per second

Exercise:

Problem: **Integrated Concepts**

A battery-operated car utilizes a 12.0 V system. Find the charge the batteries must be able to move in order to accelerate the 750 kg car from rest to 25.0 m/s, make it climb a 2.00×10^2 m high hill, and then cause it to travel at a constant 25.0 m/s by exerting a 5.00×10^2 N force for an hour.

Solution:

3.89×10^6 C

Exercise:

Problem: **Integrated Concepts**

Fusion probability is greatly enhanced when appropriate nuclei are brought close together, but mutual Coulomb repulsion must be overcome. This can be done using the kinetic energy of high-temperature gas ions or by accelerating the nuclei toward one another.

(a) Calculate the potential energy of two singly charged nuclei separated by 1.00×10^{-12} m by finding the voltage of one at that distance and multiplying by the charge of the other. (b) At what temperature will atoms of a gas have an average kinetic energy equal to this needed electrical potential energy?

Exercise:

Problem: Unreasonable Results

(a) Find the voltage near a 10.0 cm diameter metal sphere that has 8.00 C of excess positive charge on it. (b) What is unreasonable about this result? (c) Which assumptions are responsible?

Solution:

(a) 1.44×10^{12} V

(b) This voltage is very high. A 10.0 cm diameter sphere could never maintain this voltage; it would discharge.

(c) An 8.00 C charge is more charge than can reasonably be accumulated on a sphere of that size.

Exercise:

Problem: Construct Your Own Problem

Consider a battery used to supply energy to a cellular phone. Construct a problem in which you determine the energy that must be supplied by the battery, and then calculate the amount of charge it must be able to move in order to supply this energy. Among the things to be considered are the energy needs and battery voltage. You may need to look ahead to interpret manufacturer’s battery ratings in ampere-hours as energy in joules.

Glossary

electric potential
: potential energy per unit charge

potential difference (or voltage)
: change in potential energy of a charge moved from one point to another, divided by the charge; units of potential difference are joules per coulomb, known as volt

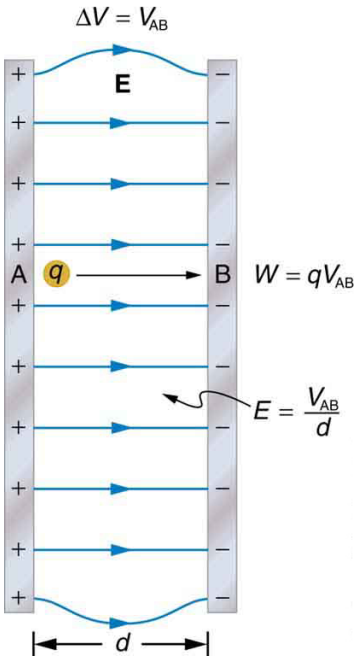
electron volt
: the energy given to a fundamental charge accelerated through a potential difference of one volt

mechanical energy
: sum of the kinetic energy and potential energy of a system; this sum is a constant

Electric Potential in a Uniform Electric Field

- Describe the relationship between voltage and electric field.
- Derive an expression for the electric potential and electric field.
- Calculate electric field strength given distance and voltage.

In the previous section, we explored the relationship between voltage and energy. In this section, we will explore the relationship between voltage and electric field. For example, a uniform electric field $\mathbf{E}$ is produced by placing a potential difference (or voltage) ΔV across two parallel metal plates, labeled A and B. (See [link].) Examining this will tell us what voltage is needed to produce a certain electric field strength; it will also reveal a more fundamental relationship between electric potential and electric field. From a physicist's point of view, either ΔV or $\mathbf{E}$ can be used to describe any charge distribution. ΔV is most closely tied to energy, whereas $\mathbf{E}$ is most closely related to force. ΔV is a **scalar** quantity and has no direction, while $\mathbf{E}$ is a **vector** quantity, having both magnitude and direction. (Note that the magnitude of the electric field strength, a scalar quantity, is represented by E below.) The relationship between ΔV and $\mathbf{E}$ is revealed by calculating the work done by the force in moving a charge from point A to point B. But, as noted in Electric Potential Energy: Potential Difference, this is complex for arbitrary charge distributions, requiring calculus. We therefore look at a uniform electric field as an interesting special case.



The relationship between V and E for parallel conducting plates is $E = V/d$. (Note that $\Delta V = V_{\text{AB}}$ in magnitude. For a charge that is moved from plate A at higher potential to plate B at lower potential, a minus sign needs to be included as follows: $-\Delta V = V_{\text{A}} - V_{\text{B}} = V_{\text{AB}}$. See the text for details.)

The work done by the electric field in [link] to move a positive charge q from A, the positive plate, higher potential, to B, the negative plate, lower potential, is

Equation:

$$W = -\Delta \mathrm{PE} = -q\Delta V.$$

The potential difference between points A and B is
Equation:

$$-\Delta V = -(V_{\mathrm{B}} - V_{\mathrm{A}}) = V_{\mathrm{A}} - V_{\mathrm{B}} = V_{\mathrm{AB}}.$$

Entering this into the expression for work yields
Equation:

$$W = qV_{\mathrm{AB}}.$$

Work is $W = Fd \cos\theta$; here $\cos\theta = 1$, since the path is parallel to the field, and so $W = Fd$. Since $F = qE$, we see that $W = qEd$. Substituting this expression for work into the previous equation gives
Equation:

$$qEd = qV_{\mathrm{AB}}.$$

The charge cancels, and so the voltage between points A and B is seen to be
Equation:

$$\left.\begin{array}{l} V_{\mathrm{AB}} = Ed \\ E = \frac{V_{\mathrm{AB}}}{d} \end{array}\right\} (\text{uniform } E \text{ - field only}),$$

where d is the distance from A to B, or the distance between the plates in [link]. Note that the above equation implies the units for electric field are volts per meter. We already know the units for electric field are newtons per coulomb; thus the following relation among units is valid:
Equation:

$$1\ \mathrm{N/C} = 1\ \mathrm{V/m}.$$

Note:
Voltage between Points A and B
Equation:

$$\left.\begin{array}{l} V_{\text{AB}} = Ed \\ E = \frac{V_{\text{AB}}}{d} \end{array}\right\} (\text{uniform } E \text{ - field only}),$$

where d is the distance from A to B, or the distance between the plates.

Example:
What Is the Highest Voltage Possible between Two Plates?
Dry air will support a maximum electric field strength of about $3.0 \times 10^6 \text{ V/m}$. Above that value, the field creates enough ionization in the air to make the air a conductor. This allows a discharge or spark that reduces the field. What, then, is the maximum voltage between two parallel conducting plates separated by 2.5 cm of dry air?
Strategy
We are given the maximum electric field E between the plates and the distance d between them. The equation $V_{\text{AB}} = Ed$ can thus be used to calculate the maximum voltage.
Solution
The potential difference or voltage between the plates is
Equation:

$$\text{V}_{\text{AB}} = Ed.$$

Entering the given values for E and d gives
Equation:

$$V_{\text{AB}} = (3.0 \times 10^6 \text{ V/m})(0.025 \text{ m}) = 7.5 \times 10^4 \text{ V}$$

or
Equation:

$$V_{\text{AB}} = 75 \text{ kV}.$$

(The answer is quoted to only two digits, since the maximum field strength is approximate.)

Discussion

One of the implications of this result is that it takes about 75 kV to make a spark jump across a 2.5 cm (1 in.) gap, or 150 kV for a 5 cm spark. This limits the voltages that can exist between conductors, perhaps on a power transmission line. A smaller voltage will cause a spark if there are points on the surface, since points create greater fields than smooth surfaces. Humid air breaks down at a lower field strength, meaning that a smaller voltage will make a spark jump through humid air. The largest voltages can be built up, say with static electricity, on dry days.



A spark chamber is used to trace the paths of high-energy particles. Ionization created by the particles as they pass through the gas between the plates allows a spark to jump. The sparks are

perpendicular to the plates, following electric field lines between them. The potential difference between adjacent plates is not high enough to cause sparks without the ionization produced by particles from accelerator experiments (or cosmic rays). (credit: Daderot, Wikimedia Commons)

Example:
Field and Force inside an Electron Gun
(a) An electron gun has parallel plates separated by 4.00 cm and gives electrons 25.0 keV of energy. What is the electric field strength between the plates? (b) What force would this field exert on a piece of plastic with a $0.500\ \mu\text{C}$ charge that gets between the plates?
Strategy
Since the voltage and plate separation are given, the electric field strength can be calculated directly from the expression $E = \frac{V_{\text{AB}}}{d}$. Once the electric field strength is known, the force on a charge is found using $\mathbf{F} = q\,\mathbf{E}$. Since the electric field is in only one direction, we can write this equation in terms of the magnitudes, $F = q\,E$.
Solution for (a)
The expression for the magnitude of the electric field between two uniform metal plates is

Equation:

$$E = \frac{V_{\text{AB}}}{d}.$$

Since the electron is a single charge and is given 25.0 keV of energy, the potential difference must be 25.0 kV. Entering this value for V_{AB} and the plate separation of 0.0400 m, we obtain
Equation:

$$E = \frac{25.0\ \text{kV}}{0.0400\ \text{m}} = 6.25 \times 10^5\ \text{V/m}.$$

Solution for (b)
The magnitude of the force on a charge in an electric field is obtained from the equation
Equation:

$$F = qE.$$

Substituting known values gives
Equation:

$$F = (0.500 \times 10^{-6}\ \text{C})(6.25 \times 10^5\ \text{V/m}) = 0.313\ \text{N}.$$

Discussion
Note that the units are newtons, since $1\ \text{V/m} = 1\ \text{N/C}$. The force on the charge is the same no matter where the charge is located between the plates. This is because the electric field is uniform between the plates.

In more general situations, regardless of whether the electric field is uniform, it points in the direction of decreasing potential, because the force on a positive charge is in the direction of $\mathbf{E}$ and also in the direction of lower potential V. Furthermore, the magnitude of $\mathbf{E}$ equals the rate of decrease of V with distance. The faster V decreases over distance, the

greater the electric field. In equation form, the general relationship between voltage and electric field is
Equation:

$$E = -\frac{\Delta V}{\Delta s},$$

where Δs is the distance over which the change in potential, ΔV, takes place. The minus sign tells us that $\mathbf{E}$ points in the direction of decreasing potential. The electric field is said to be the *gradient* (as in grade or slope) of the electric potential.

Note:
Relationship between Voltage and Electric Field
In equation form, the general relationship between voltage and electric field is
Equation:

$$E = -\frac{\Delta V}{\Delta s},$$

where Δs is the distance over which the change in potential, ΔV, takes place. The minus sign tells us that $\mathbf{E}$ points in the direction of decreasing potential. The electric field is said to be the *gradient* (as in grade or slope) of the electric potential.

For continually changing potentials, ΔV and Δs become infinitesimals and differential calculus must be employed to determine the electric field.

Section Summary

- The voltage between points A and B is
  Equation:

$$\left.\begin{array}{l} V_{\text{AB}} = Ed \\ E = \frac{V_{\text{AB}}}{d} \end{array}\right\} (\text{uniform } E \text{ - field only}),$$

where d is the distance from A to B, or the distance between the plates.

- In equation form, the general relationship between voltage and electric field is
  Equation:

$$E = -\frac{\Delta V}{\Delta s},$$

where Δs is the distance over which the change in potential, ΔV, takes place. The minus sign tells us that $\mathbf{E}$ points in the direction of decreasing potential.) The electric field is said to be the *gradient* (as in grade or slope) of the electric potential.

Conceptual Questions

Exercise:

Problem:

Discuss how potential difference and electric field strength are related. Give an example.

Exercise:

Problem:

What is the strength of the electric field in a region where the electric potential is constant?

Exercise:

Problem:

Will a negative charge, initially at rest, move toward higher or lower potential? Explain why.

Problems & Exercises

Exercise:

Problem:

Show that units of V/m and N/C for electric field strength are indeed equivalent.

Exercise:

Problem:

What is the strength of the electric field between two parallel conducting plates separated by 1.00 cm and having a potential difference (voltage) between them of 1.50×10^4 V?

Exercise:

Problem:

The electric field strength between two parallel conducting plates separated by 4.00 cm is 7.50×10^4 V/m. (a) What is the potential difference between the plates? (b) The plate with the lowest potential is taken to be at zero volts. What is the potential 1.00 cm from that plate (and 3.00 cm from the other)?

Solution:

(a) 3.00 kV

(b) 750 V

Exercise:

Problem:

How far apart are two conducting plates that have an electric field strength of 4.50×10^3 V/m between them, if their potential difference is 15.0 kV?

Exercise:

Problem:

(a) Will the electric field strength between two parallel conducting plates exceed the breakdown strength for air ($3.0 \times 10^6\ \mathrm{V/m}$) if the plates are separated by 2.00 mm and a potential difference of $5.0 \times 10^3\ \mathrm{V}$ is applied? (b) How close together can the plates be with this applied voltage?

Solution:

(a) No. The electric field strength between the plates is $2.5 \times 10^6\ \mathrm{V/m}$, which is lower than the breakdown strength for air ($3.0 \times 10^6\ \mathrm{V/m}$).

(b) 1.7 mm

Exercise:

Problem:

The voltage across a membrane forming a cell wall is 80.0 mV and the membrane is 9.00 nm thick. What is the electric field strength? (The value is surprisingly large, but correct. Membranes are discussed in Capacitors and Dielectrics and Nerve Conduction—Electrocardiograms.) You may assume a uniform electric field.

Exercise:

Problem:

Membrane walls of living cells have surprisingly large electric fields across them due to separation of ions. (Membranes are discussed in some detail in Nerve Conduction—Electrocardiograms.) What is the voltage across an 8.00 nm–thick membrane if the electric field strength across it is 5.50 MV/m? You may assume a uniform electric field.

Solution:

44.0 mV

Exercise:

Problem:

Two parallel conducting plates are separated by 10.0 cm, and one of them is taken to be at zero volts. (a) What is the electric field strength between them, if the potential 8.00 cm from the zero volt plate (and 2.00 cm from the other) is 450 V? (b) What is the voltage between the plates?

Exercise:

Problem:

Find the maximum potential difference between two parallel conducting plates separated by 0.500 cm of air, given the maximum sustainable electric field strength in air to be $3.0 \times 10^6\ \mathrm{V/m}$.

Solution:

$15\ \mathrm{kV}$

Exercise:

Problem:

A doubly charged ion is accelerated to an energy of 32.0 keV by the electric field between two parallel conducting plates separated by 2.00 cm. What is the electric field strength between the plates?

Exercise:

Problem:

An electron is to be accelerated in a uniform electric field having a strength of $2.00 \times 10^6\ \mathrm{V/m}$. (a) What energy in keV is given to the electron if it is accelerated through 0.400 m? (b) Over what distance would it have to be accelerated to increase its energy by 50.0 GeV?

Solution:

(a) $800\ \mathrm{KeV}$

(b) 25.0 km

Glossary

scalar
: physical quantity with magnitude but no direction

vector
: physical quantity with both magnitude and direction

Electrical Potential Due to a Point Charge

- Explain point charges and express the equation for electric potential of a point charge.
- Distinguish between electric potential and electric field.
- Determine the electric potential of a point charge given charge and distance.

Point charges, such as electrons, are among the fundamental building blocks of matter. Furthermore, spherical charge distributions (like on a metal sphere) create external electric fields exactly like a point charge. The electric potential due to a point charge is, thus, a case we need to consider. Using calculus to find the work needed to move a test charge *q* from a large distance away to a distance of *r* from a point charge Q, and noting the connection between work and potential $(W = -q\Delta V)$, it can be shown that the *electric potential* V *of a point charge* is
Equation:

$$V = \frac{kQ}{r} \text{ (Point Charge)},$$

where *k* is a constant equal to $9.0 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2$.

Note:
Electric Potential V of a Point Charge
The electric potential V of a point charge is given by
Equation:

$$V = \frac{kQ}{r} \text{ (Point Charge)}.$$

The potential at infinity is chosen to be zero. Thus V for a point charge decreases with distance, whereas $\mathbf{E}$ for a point charge decreases with distance squared:

Equation:

$$E = \frac{F}{q} = \frac{kQ}{r^2}.$$

Recall that the electric potential V is a scalar and has no direction, whereas the electric field $\mathbf{E}$ is a vector. To find the voltage due to a combination of point charges, you add the individual voltages as numbers. To find the total electric field, you must add the individual fields as *vectors*, taking magnitude and direction into account. This is consistent with the fact that V is closely associated with energy, a scalar, whereas $\mathbf{E}$ is closely associated with force, a vector.

Example:

What Voltage Is Produced by a Small Charge on a Metal Sphere?

Charges in static electricity are typically in the nanocoulomb (nC) to microcoulomb $(\mu\mathrm{C})$ range. What is the voltage 5.00 cm away from the center of a 1-cm diameter metal sphere that has a $-3.00\ \mathrm{nC}$ static charge?

Strategy

As we have discussed in Electric Charge and Electric Field, charge on a metal sphere spreads out uniformly and produces a field like that of a point charge located at its center. Thus we can find the voltage using the equation $V = \mathrm{kQ}/r$.

Solution

Entering known values into the expression for the potential of a point charge, we obtain

Equation:

$$\begin{aligned} V &= k\frac{Q}{r} \\ &= \left(8.99 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2\right)\left(\frac{-3.00 \times 10^{-9} \text{ C}}{5.00 \times 10^{-2} \text{ m}}\right) \\ &= -539 \text{ V}. \end{aligned}$$

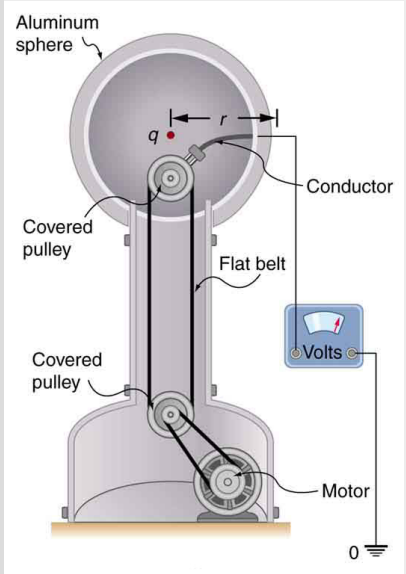
Discussion

The negative value for voltage means a positive charge would be attracted from a larger distance, since the potential is lower (more negative) than at larger distances. Conversely, a negative charge would be repelled, as expected.

Example:

What Is the Excess Charge on a Van de Graaff Generator

A demonstration Van de Graaff generator has a 25.0 cm diameter metal sphere that produces a voltage of 100 kV near its surface. (See [link].) What excess charge resides on the sphere? (Assume that each numerical value here is shown with three significant figures.)



The voltage of this demonstration Van de Graaff generator is measured

between the charged sphere and ground. Earth's potential is taken to be zero as a reference. The potential of the charged conducting sphere is the same as that of an equal point charge at its center.

Strategy

The potential on the surface will be the same as that of a point charge at the center of the sphere, 12.5 cm away. (The radius of the sphere is 12.5 cm.) We can thus determine the excess charge using the equation

Equation:

$$V = \frac{\mathrm{kQ}}{r}.$$

Solution

Solving for Q and entering known values gives

Equation:

$$\begin{aligned} Q &= \frac{\mathrm{rV}}{k} \\ &= \frac{(0.125\ \mathrm{m})\left(100\times10^{3}\ \mathrm{V}\right)}{8.99\times10^{9}\ \mathrm{N\cdot m^2/C^2}} \\ &= 1.39\times10^{-6}\ \mathrm{C} = 1.39\ \mu\mathrm{C}. \end{aligned}$$

Discussion

This is a relatively small charge, but it produces a rather large voltage. We have another indication here that it is difficult to store isolated charges.

The voltages in both of these examples could be measured with a meter that compares the measured potential with ground potential. Ground potential is often taken to be zero (instead of taking the potential at infinity to be zero). It is the potential difference between two points that is of importance, and very often there is a tacit assumption that some reference point, such as Earth or a very distant point, is at zero potential. As noted in Electric Potential Energy: Potential Difference, this is analogous to taking sea level as $h = 0$ when considering gravitational potential energy, $\mathrm{PE_g = mgh}$.

Section Summary

- Electric potential of a point charge is $V = \mathrm{kQ}/r$.
- Electric potential is a scalar, and electric field is a vector. Addition of voltages as numbers gives the voltage due to a combination of point charges, whereas addition of individual fields as vectors gives the total electric field.

Conceptual Questions

Exercise:

Problem:

In what region of space is the potential due to a uniformly charged sphere the same as that of a point charge? In what region does it differ from that of a point charge?

Exercise:

Problem:

Can the potential of a non-uniformly charged sphere be the same as that of a point charge? Explain.

Problems & Exercises

Exercise:

Problem:

A 0.500 cm diameter plastic sphere, used in a static electricity demonstration, has a uniformly distributed 40.0 pC charge on its surface. What is the potential near its surface?

Solution:

144 V

Exercise:

Problem:

What is the potential 0.530×10^{-10} m from a proton (the average distance between the proton and electron in a hydrogen atom)?

Exercise:

Problem:

(a) A sphere has a surface uniformly charged with 1.00 C. At what distance from its center is the potential 5.00 MV? (b) What does your answer imply about the practical aspect of isolating such a large charge?

Solution:

(a) 1.80 km

(b) A charge of 1 C is a very large amount of charge; a sphere of radius 1.80 km is not practical.

Exercise:

Problem:

How far from a 1.00 μC point charge will the potential be 100 V? At what distance will it be 2.00×10^2 V?

Exercise:

Problem:

What are the sign and magnitude of a point charge that produces a potential of $-2.00\ \text{V}$ at a distance of 1.00 mm?

Solution:

$-2.22 \times 10^{-13}\ \text{C}$

Exercise:

Problem:

If the potential due to a point charge is $5.00 \times 10^{2}\ \text{V}$ at a distance of 15.0 m, what are the sign and magnitude of the charge?

Exercise:

Problem:

In nuclear fission, a nucleus splits roughly in half. (a) What is the potential $2.00 \times 10^{-14}\ \text{m}$ from a fragment that has 46 protons in it? (b) What is the potential energy in MeV of a similarly charged fragment at this distance?

Solution:

(a) $3.31 \times 10^{6}\ \text{V}$

(b) 152 MeV

Exercise:

Problem:

A research Van de Graaff generator has a 2.00-m-diameter metal sphere with a charge of 5.00 mC on it. (a) What is the potential near its surface? (b) At what distance from its center is the potential 1.00 MV? (c) An oxygen atom with three missing electrons is released near the Van de Graaff generator. What is its energy in MeV at this distance?

Exercise:

Problem:

An electrostatic paint sprayer has a 0.200-m-diameter metal sphere at a potential of 25.0 kV that repels paint droplets onto a grounded object. (a) What charge is on the sphere? (b) What charge must a 0.100-mg drop of paint have to arrive at the object with a speed of 10.0 m/s?

Solution:

(a) $2.78 \times 10^{-7}\ \text{C}$

(b) $2.00 \times 10^{-10}\ \text{C}$

Exercise:

Problem:

In one of the classic nuclear physics experiments at the beginning of the 20th century, an alpha particle was accelerated toward a gold nucleus, and its path was substantially deflected by the Coulomb interaction. If the energy of the doubly charged alpha nucleus was 5.00 MeV, how close to the gold nucleus (79 protons) could it come before being deflected?

Exercise:

Problem:

(a) What is the potential between two points situated 10 cm and 20 cm from a $3.0\ \mu\text{C}$ point charge? (b) To what location should the point at 20 cm be moved to increase this potential difference by a factor of two?

Exercise:

Problem: Unreasonable Results

(a) What is the final speed of an electron accelerated from rest through a voltage of 25.0 MV by a negatively charged Van de Graaff terminal?

(b) What is unreasonable about this result?

(c) Which assumptions are responsible?

Solution:

(a) $2.96 \times 10^9 \text{ m/s}$

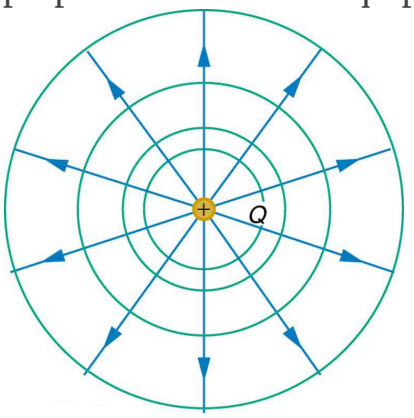
(b) This velocity is far too great. It is faster than the speed of light.

(c) The assumption that the speed of the electron is far less than that of light and that the problem does not require a relativistic treatment produces an answer greater than the speed of light.

Equipotential Lines

- Explain equipotential lines and equipotential surfaces.
- Describe the action of grounding an electrical appliance.
- Compare electric field and equipotential lines.

We can represent electric potentials (voltages) pictorially, just as we drew pictures to illustrate electric fields. Of course, the two are related. Consider [link], which shows an isolated positive point charge and its electric field lines. Electric field lines radiate out from a positive charge and terminate on negative charges. While we use blue arrows to represent the magnitude and direction of the electric field, we use green lines to represent places where the electric potential is constant. These are called **equipotential lines** in two dimensions, or *equipotential surfaces* in three dimensions. The term *equipotential* is also used as a noun, referring to an equipotential line or surface. The potential for a point charge is the same anywhere on an imaginary sphere of radius r surrounding the charge. This is true since the potential for a point charge is given by $V = \mathrm{kQ}/r$ and, thus, has the same value at any point that is a given distance r from the charge. An equipotential sphere is a circle in the two-dimensional view of [link]. Since the electric field lines point radially away from the charge, they are perpendicular to the equipotential lines.



An isolated point charge Q with its electric field lines in blue and equipotential lines

in green. The potential is the same along each equipotential line, meaning that no work is required to move a charge anywhere along one of those lines. Work is needed to move a charge from one equipotential line to another. Equipotential lines are perpendicular to electric field lines in every case.

It is important to note that *equipotential lines are always perpendicular to electric field lines*. No work is required to move a charge along an equipotential, since $\Delta V = 0$. Thus the work is

Equation:

$$W = -\Delta \text{ PE} = -q\Delta V = 0.$$

Work is zero if force is perpendicular to motion. Force is in the same direction as $\mathbf{E}$, so that motion along an equipotential must be perpendicular to $\mathbf{E}$. More precisely, work is related to the electric field by

Equation:

$$W = Fd \cos \theta = qEd \cos \theta = 0.$$

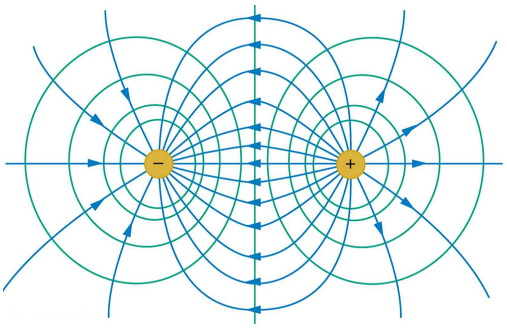
Note that in the above equation, E and F symbolize the magnitudes of the electric field strength and force, respectively. Neither q nor $\mathbf{E}$ nor d is zero, and so cos θ must be 0, meaning θ must be 90º. In other words, motion along an equipotential is perpendicular to $\mathbf{E}$.

One of the rules for static electric fields and conductors is that the electric field must be perpendicular to the surface of any conductor. This implies that a *conductor is an equipotential surface in static situations*. There can be no voltage difference across the surface of a conductor, or charges will flow. One of the uses of this fact is that a conductor can be fixed at zero volts by connecting it to the earth with a good conductor—a process called **grounding**. Grounding can be a useful safety tool. For example, grounding the metal case of an electrical appliance ensures that it is at zero volts relative to the earth.

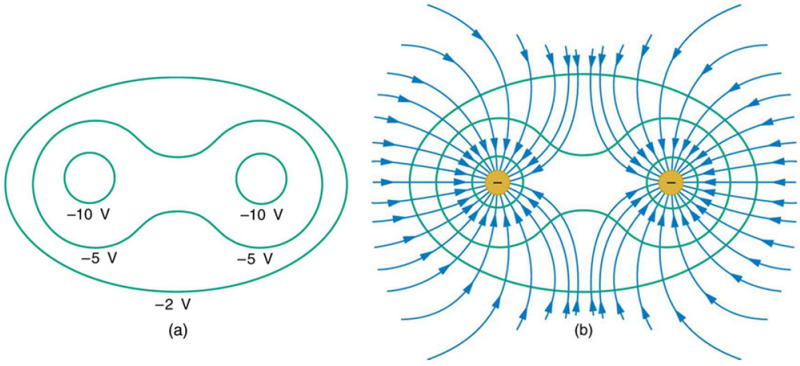
Note:
Grounding
A conductor can be fixed at zero volts by connecting it to the earth with a good conductor—a process called grounding.

Because a conductor is an equipotential, it can replace any equipotential surface. For example, in [link] a charged spherical conductor can replace the point charge, and the electric field and potential surfaces outside of it will be unchanged, confirming the contention that a spherical charge distribution is equivalent to a point charge at its center.

[link] shows the electric field and equipotential lines for two equal and opposite charges. Given the electric field lines, the equipotential lines can be drawn simply by making them perpendicular to the electric field lines. Conversely, given the equipotential lines, as in [link](a), the electric field lines can be drawn by making them perpendicular to the equipotentials, as in [link](b).



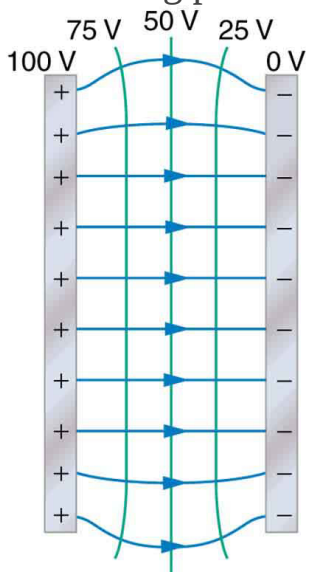
The electric field lines and equipotential lines for two equal but opposite charges. The equipotential lines can be drawn by making them perpendicular to the electric field lines, if those are known. Note that the potential is greatest (most positive) near the positive charge and least (most negative) near the negative charge.



(a) These equipotential lines might be measured with a voltmeter in a laboratory experiment. (b) The corresponding electric field lines are found by drawing them

perpendicular to the equipotentials. Note that these fields are consistent with two equal negative charges.

One of the most important cases is that of the familiar parallel conducting plates shown in [link]. Between the plates, the equipotentials are evenly spaced and parallel. The same field could be maintained by placing conducting plates at the equipotential lines at the potentials shown.



The electric field and equipotential lines between two metal plates.

An important application of electric fields and equipotential lines involves the heart. The heart relies on electrical signals to maintain its rhythm. The movement of electrical signals causes the chambers of the heart to contract and relax. When a person has a heart attack, the movement of these electrical signals may be disturbed. An artificial pacemaker and a

defibrillator can be used to initiate the rhythm of electrical signals. The equipotential lines around the heart, the thoracic region, and the axis of the heart are useful ways of monitoring the structure and functions of the heart. An electrocardiogram (ECG) measures the small electric signals being generated during the activity of the heart. More about the relationship between electric fields and the heart is discussed in Energy Stored in Capacitors.

Note:
PhET Explorations: Charges and Fields
Move point charges around on the playing field and then view the electric field, voltages, equipotential lines, and more. It's colorful, it's dynamic, it's free.
https://phet.colorado.edu/sims/html/charges-and-fields/latest/charges-and-fields_en.html

Section Summary

- An equipotential line is a line along which the electric potential is constant.
- An equipotential surface is a three-dimensional version of equipotential lines.
- Equipotential lines are always perpendicular to electric field lines.
- The process by which a conductor can be fixed at zero volts by connecting it to the earth with a good conductor is called grounding.

Conceptual Questions

Exercise:

Problem:

What is an equipotential line? What is an equipotential surface?

Exercise:

Problem:

Explain in your own words why equipotential lines and surfaces must be perpendicular to electric field lines.

Exercise:

Problem: Can different equipotential lines cross? Explain.

Problems & Exercises

Exercise:

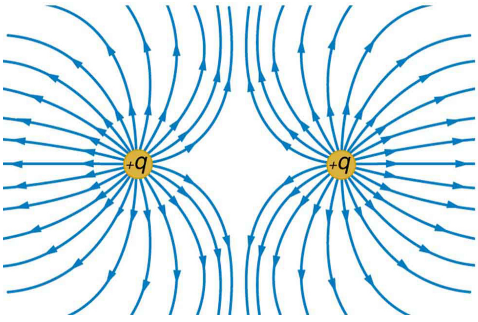
Problem:

(a) Sketch the equipotential lines near a point charge + q. Indicate the direction of increasing potential. (b) Do the same for a point charge $-3\ q$.

Exercise:

Problem:

Sketch the equipotential lines for the two equal positive charges shown in [link]. Indicate the direction of increasing potential.



The electric field near two equal positive charges is directed away from each of the charges.

Exercise:

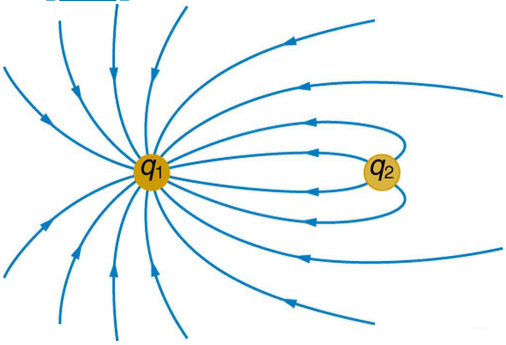
Problem:

[link] shows the electric field lines near two charges q_1 and q_2, the first having a magnitude four times that of the second. Sketch the equipotential lines for these two charges, and indicate the direction of increasing potential.

Exercise:

Problem:

Sketch the equipotential lines a long distance from the charges shown in [link]. Indicate the direction of increasing potential.



The electric field near two charges.

Exercise:

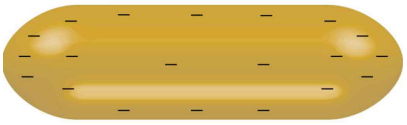
Problem:

Sketch the equipotential lines in the vicinity of two opposite charges, where the negative charge is three times as great in magnitude as the positive. See [link] for a similar situation. Indicate the direction of increasing potential.

Exercise:

Problem:

Sketch the equipotential lines in the vicinity of the negatively charged conductor in [link]. How will these equipotentials look a long distance from the object?

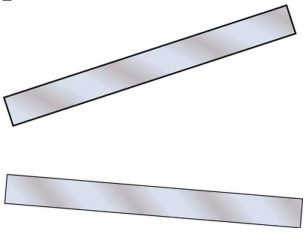


A negatively charged conductor.

Exercise:

Problem:

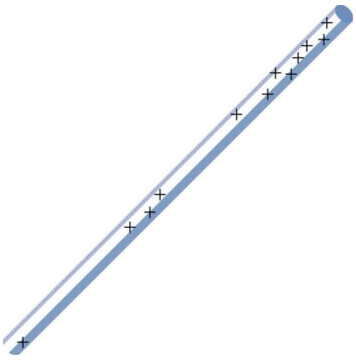
Sketch the equipotential lines surrounding the two conducting plates shown in [link], given the top plate is positive and the bottom plate has an equal amount of negative charge. Be certain to indicate the distribution of charge on the plates. Is the field strongest where the plates are closest? Why should it be?



Exercise:

Problem:

(a) Sketch the electric field lines in the vicinity of the charged insulator in [link]. Note its non-uniform charge distribution. (b) Sketch equipotential lines surrounding the insulator. Indicate the direction of increasing potential.



A charged insulating rod such as might be used in a classroom demonstration.

Exercise:

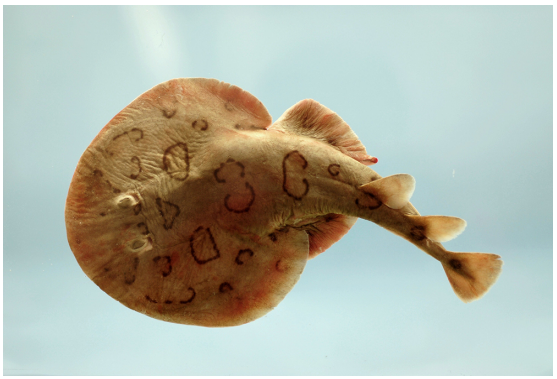
Problem:

The naturally occurring charge on the ground on a fine day out in the open country is $-1.00\ \mathrm{nC/m^2}$. (a) What is the electric field relative to ground at a height of 3.00 m? (b) Calculate the electric potential at this height. (c) Sketch electric field and equipotential lines for this scenario.

Exercise:

Problem:

The lesser electric ray (*Narcine bancroftii*) maintains an incredible charge on its head and a charge equal in magnitude but opposite in sign on its tail ([link]). (a) Sketch the equipotential lines surrounding the ray. (b) Sketch the equipotentials when the ray is near a ship with a conducting surface. (c) How could this charge distribution be of use to the ray?



Lesser electric ray (*Narcine bancroftii*) (credit: National Oceanic and Atmospheric Administration, NOAA's Fisheries Collection).

Glossary

equipotential line
: a line along which the electric potential is constant

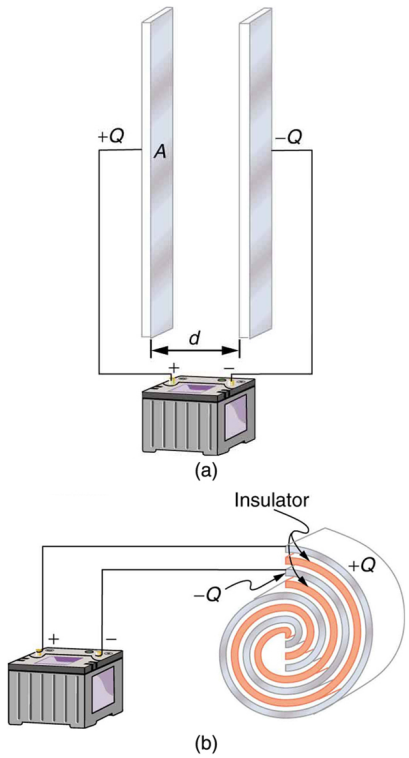
grounding
: fixing a conductor at zero volts by connecting it to the earth or ground

Capacitors and Dielectrics

- Describe the action of a capacitor and define capacitance.
- Explain parallel plate capacitors and their capacitances.
- Discuss the process of increasing the capacitance of a dielectric.
- Determine capacitance given charge and voltage.

A **capacitor** is a device used to store electric charge. Capacitors have applications ranging from filtering static out of radio reception to energy storage in heart defibrillators. Typically, commercial capacitors have two conducting parts close to one another, but not touching, such as those in [link]. (Most of the time an insulator is used between the two plates to provide separation—see the discussion on dielectrics below.) When battery terminals are connected to an initially uncharged capacitor, equal amounts of positive and negative charge, $+Q$ and $-Q$, are separated into its two plates. The capacitor remains neutral overall, but we refer to it as storing a charge Q in this circumstance.

Note:
Capacitor
A capacitor is a device used to store electric charge.

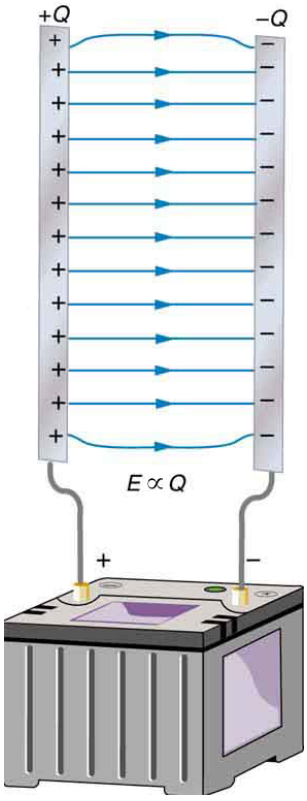


Both capacitors shown here were initially uncharged before being connected to a battery. They now have separated charges of $+Q$ and $-Q$ on their two halves. (a) A parallel plate capacitor. (b) A rolled capacitor with an insulating material between its two conducting sheets.

The amount of charge Q a *capacitor* can store depends on two major factors—the voltage applied and the capacitor's physical characteristics, such as its size.

Note:
The Amount of Charge Q a Capacitor Can Store
The amount of charge Q a *capacitor* can store depends on two major factors—the voltage applied and the capacitor's physical characteristics, such as its size.

A system composed of two identical, parallel conducting plates separated by a distance, as in [link], is called a **parallel plate capacitor**. It is easy to see the relationship between the voltage and the stored charge for a parallel plate capacitor, as shown in [link]. Each electric field line starts on an individual positive charge and ends on a negative one, so that there will be more field lines if there is more charge. (Drawing a single field line per charge is a convenience, only. We can draw many field lines for each charge, but the total number is proportional to the number of charges.) The electric field strength is, thus, directly proportional to Q.



Electric field lines in this parallel plate capacitor, as always, start on positive charges and end on negative charges. Since the electric field strength is proportional to the density of field lines, it is also proportional to the amount

of charge on
the capacitor.

The field is proportional to the charge:
Equation:

$$E \propto Q,$$

where the symbol $\propto$ means "proportional to." From the discussion in Electric Potential in a Uniform Electric Field, we know that the voltage across parallel plates is $V = \text{Ed}$. Thus,
Equation:

$$V \propto E.$$

It follows, then, that $V \propto Q$, and conversely,
Equation:

$$Q \propto V.$$

This is true in general: The greater the voltage applied to any capacitor, the greater the charge stored in it.

Different capacitors will store different amounts of charge for the same applied voltage, depending on their physical characteristics. We define their **capacitance** C to be such that the charge Q stored in a capacitor is proportional to C. The charge stored in a capacitor is given by
Equation:

$$Q = \text{CV}.$$

This equation expresses the two major factors affecting the amount of charge stored. Those factors are the physical characteristics of the capacitor,

C, and the voltage, V. Rearranging the equation, we see that *capacitance C is the amount of charge stored per volt*, or
Equation:

$$C = \frac{Q}{V}.$$

Note:
Capacitance
Capacitance C is the amount of charge stored per volt, or
Equation:

$$C = \frac{Q}{V}.$$

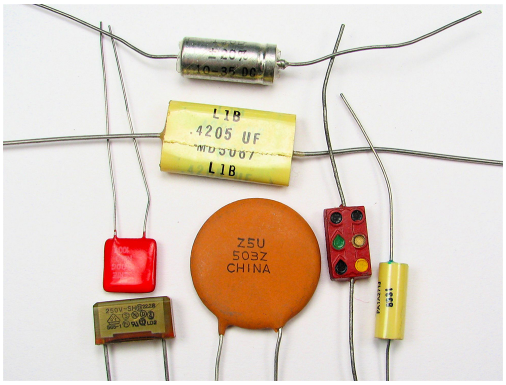
The unit of capacitance is the farad (F), named for Michael Faraday (1791–1867), an English scientist who contributed to the fields of electromagnetism and electrochemistry. Since capacitance is charge per unit voltage, we see that a farad is a coulomb per volt, or
Equation:

$$1\ \text{F} = \frac{1\ \text{C}}{1\ \text{V}}.$$

A 1-farad capacitor would be able to store 1 coulomb (a very large amount of charge) with the application of only 1 volt. One farad is, thus, a very large capacitance. Typical capacitors range from fractions of a picofarad $\left(1\ \text{pF} = 10^{-12}\ \text{F}\right)$ to millifarads $\left(1\ \text{mF} = 10^{-3}\ \text{F}\right)$.

[link] shows some common capacitors. Capacitors are primarily made of ceramic, glass, or plastic, depending upon purpose and size. Insulating

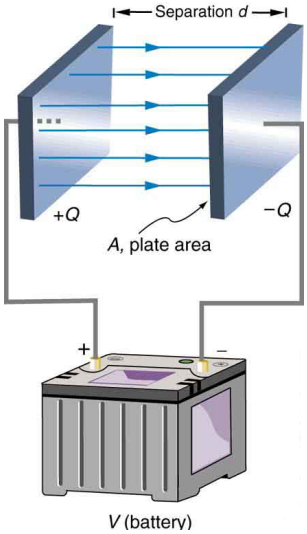
materials, called dielectrics, are commonly used in their construction, as discussed below.



Some typical capacitors. Size and value of capacitance are not necessarily related. (credit: Windell Oskay)

Parallel Plate Capacitor

The parallel plate capacitor shown in [link] has two identical conducting plates, each having a surface area A, separated by a distance d (with no material between the plates). When a voltage V is applied to the capacitor, it stores a charge Q, as shown. We can see how its capacitance depends on A and d by considering the characteristics of the Coulomb force. We know that like charges repel, unlike charges attract, and the force between charges decreases with distance. So it seems quite reasonable that the bigger the plates are, the more charge they can store—because the charges can spread out more. Thus C should be greater for larger A. Similarly, the closer the plates are together, the greater the attraction of the opposite charges on them. So C should be greater for smaller d.



Parallel plate capacitor with plates separated by a distance d. Each plate has an area A.

It can be shown that for a parallel plate capacitor there are only two factors (A and d) that affect its capacitance C. The capacitance of a parallel plate capacitor in equation form is given by

Equation:

$$C = \varepsilon_0 \frac{A}{d}.$$

Note:
Capacitance of a Parallel Plate Capacitor
Equation:

$$C = \varepsilon_0 \frac{A}{d}$$

A is the area of one plate in square meters, and d is the distance between the plates in meters. The constant ε_0 is the permittivity of free space; its numerical value in SI units is $\varepsilon_0 = 8.85 \times 10^{-12}\ \mathrm{F/m}$. The units of F/m are equivalent to $\mathrm{C^2/N \cdot m^2}$. The small numerical value of ε_0 is related to the large size of the farad. A parallel plate capacitor must have a large area to have a capacitance approaching a farad. (Note that the above equation is valid when the parallel plates are separated by air or free space. When another material is placed between the plates, the equation is modified, as discussed below.)

Example:
Capacitance and Charge Stored in a Parallel Plate Capacitor
(a) What is the capacitance of a parallel plate capacitor with metal plates, each of area $1.00\ \mathrm{m^2}$, separated by 1.00 mm? (b) What charge is stored in this capacitor if a voltage of $3.00 \times 10^3\ \mathrm{V}$ is applied to it?
Strategy
Finding the capacitance C is a straightforward application of the equation $C = \varepsilon_0 A/d$. Once C is found, the charge stored can be found using the equation $Q = \mathrm{CV}$.
Solution for (a)
Entering the given values into the equation for the capacitance of a parallel plate capacitor yields
Equation:

$$\begin{aligned} C &= \varepsilon_0 \tfrac{A}{d} = \left(8.85 \times 10^{-12}\ \tfrac{\mathrm{F}}{\mathrm{m}}\right) \tfrac{1.00\ \mathrm{m^2}}{1.00 \times 10^{-3}\ \mathrm{m}} \\ &= 8.85 \times 10^{-9}\ \mathrm{F} = 8.85\ \mathrm{nF}. \end{aligned}$$

Discussion for (a)

This small value for the capacitance indicates how difficult it is to make a device with a large capacitance. Special techniques help, such as using very large area thin foils placed close together.

Solution for (b)

The charge stored in any capacitor is given by the equation $Q = \mathrm{CV}$. Entering the known values into this equation gives

Equation:

$$\begin{aligned} Q &= CV = \left(8.85 \times 10^{-9}\ \mathrm{F}\right)\left(3.00 \times 10^{3}\ \mathrm{V}\right) \\ &= 26.6\ \mu\mathrm{C}. \end{aligned}$$

Discussion for (b)

This charge is only slightly greater than those found in typical static electricity. Since air breaks down at about $3.00 \times 10^{6}\ \mathrm{V/m}$, more charge cannot be stored on this capacitor by increasing the voltage.

Another interesting biological example dealing with electric potential is found in the cell's plasma membrane. The membrane sets a cell off from its surroundings and also allows ions to selectively pass in and out of the cell. There is a potential difference across the membrane of about $-70\ \mathrm{mV}$. This is due to the mainly negatively charged ions in the cell and the predominance of positively charged sodium (Na^{+}) ions outside. Things change when a nerve cell is stimulated. Na^{+} ions are allowed to pass through the membrane into the cell, producing a positive membrane potential—the nerve signal. The cell membrane is about 7 to 10 nm thick. An approximate value of the electric field across it is given by

Equation:

$$E = \frac{V}{d} = \frac{-70 \times 10^{-3}\ \mathrm{V}}{8 \times 10^{-9}\ \mathrm{m}} = -9 \times 10^{6}\ \mathrm{V/m}.$$

This electric field is enough to cause a breakdown in air.

Dielectric

The previous example highlights the difficulty of storing a large amount of charge in capacitors. If d is made smaller to produce a larger capacitance, then the maximum voltage must be reduced proportionally to avoid breakdown (since $E = V/d$). An important solution to this difficulty is to put an insulating material, called a **dielectric**, between the plates of a capacitor and allow d to be as small as possible. Not only does the smaller d make the capacitance greater, but many insulators can withstand greater electric fields than air before breaking down.

There is another benefit to using a dielectric in a capacitor. Depending on the material used, the capacitance is greater than that given by the equation $C = \varepsilon_0 \frac{A}{d}$ by a factor κ, called the *dielectric constant*. A parallel plate capacitor with a dielectric between its plates has a capacitance given by
Equation:

$$C = \kappa \varepsilon_0 \frac{A}{d} \text{ (parallel plate capacitor with dielectric).}$$

Values of the dielectric constant κ for various materials are given in [link]. Note that κ for vacuum is exactly 1, and so the above equation is valid in that case, too. If a dielectric is used, perhaps by placing Teflon between the plates of the capacitor in [link], then the capacitance is greater by the factor κ, which for Teflon is 2.1.

Note:
Take-Home Experiment: Building a Capacitor
How large a capacitor can you make using a chewing gum wrapper? The plates will be the aluminum foil, and the separation (dielectric) in between will be the paper.

Material	Dielectric constant κ	Dielectric strength (V/m)
Vacuum	1.00000	—
Air	1.00059	3×10^6
Bakelite	4.9	24×10^6
Fused quartz	3.78	8×10^6
Neoprene rubber	6.7	12×10^6
Nylon	3.4	14×10^6
Paper	3.7	16×10^6
Polystyrene	2.56	24×10^6
Pyrex glass	5.6	14×10^6
Silicon oil	2.5	15×10^6
Strontium titanate	233	8×10^6
Teflon	2.1	60×10^6
Water	80	—

Dielectric Constants and Dielectric Strengths for Various Materials at 20ºC

Note also that the dielectric constant for air is very close to 1, so that air-filled capacitors act much like those with vacuum between their plates *except* that the air can become conductive if the electric field strength

becomes too great. (Recall that $E = V/d$ for a parallel plate capacitor.) Also shown in [link] are maximum electric field strengths in V/m, called **dielectric strengths**, for several materials. These are the fields above which the material begins to break down and conduct. The dielectric strength imposes a limit on the voltage that can be applied for a given plate separation. For instance, in [link], the separation is 1.00 mm, and so the voltage limit for air is

Equation:

$$\begin{aligned} V &= E \cdot d \\ &= (3 \times 10^6 \text{ V/m})(1.00 \times 10^{-3} \text{ m}) \\ &= 3000 \text{ V}. \end{aligned}$$

However, the limit for a 1.00 mm separation filled with Teflon is 60,000 V, since the dielectric strength of Teflon is 60×10^6 V/m. So the same capacitor filled with Teflon has a greater capacitance and can be subjected to a much greater voltage. Using the capacitance we calculated in the above example for the air-filled parallel plate capacitor, we find that the Teflon-filled capacitor can store a maximum charge of

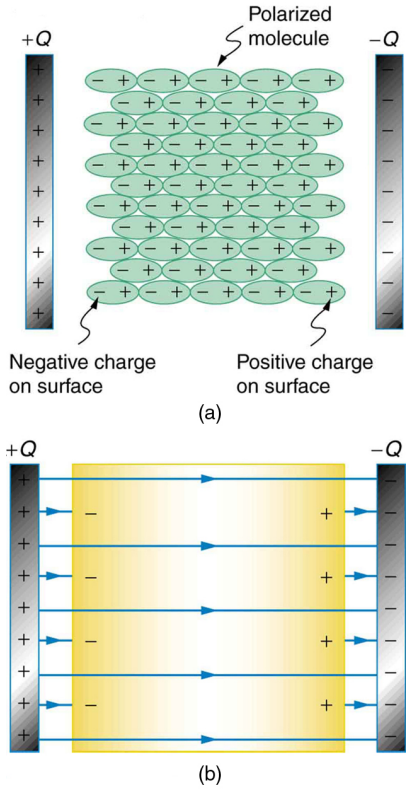
Equation:

$$\begin{aligned} Q &= CV \\ &= \kappa C_{\text{air}} V \\ &= (2.1)(8.85 \text{ nF})(6.0 \times 10^4 \text{ V}) \\ &= 1.1 \text{ mC}. \end{aligned}$$

This is 42 times the charge of the same air-filled capacitor.

Note:
Dielectric Strength
The maximum electric field strength above which an insulating material begins to break down and conduct is called its dielectric strength.

Microscopically, how does a dielectric increase capacitance? Polarization of the insulator is responsible. The more easily it is polarized, the greater its dielectric constant κ. Water, for example, is a **polar molecule** because one end of the molecule has a slight positive charge and the other end has a slight negative charge. The polarity of water causes it to have a relatively large dielectric constant of 80. The effect of polarization can be best explained in terms of the characteristics of the Coulomb force. [link] shows the separation of charge schematically in the molecules of a dielectric material placed between the charged plates of a capacitor. The Coulomb force between the closest ends of the molecules and the charge on the plates is attractive and very strong, since they are very close together. This attracts more charge onto the plates than if the space were empty and the opposite charges were a distance d away.



(a) The molecules in the insulating material between

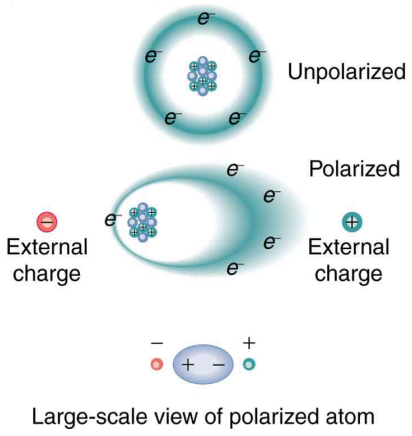
the plates of a capacitor are polarized by the charged plates. This produces a layer of opposite charge on the surface of the dielectric that attracts more charge onto the plate, increasing its capacitance. (b) The dielectric reduces the electric field strength inside the capacitor, resulting in a smaller voltage between the plates for the same charge. The capacitor stores the same charge for a smaller voltage, implying that it has a larger capacitance because of the dielectric.

Another way to understand how a dielectric increases capacitance is to consider its effect on the electric field inside the capacitor. [link](b) shows the electric field lines with a dielectric in place. Since the field lines end on charges in the dielectric, there are fewer of them going from one side of the capacitor to the other. So the electric field strength is less than if there were

a vacuum between the plates, even though the same charge is on the plates. The voltage between the plates is $V = \mathrm{Ed}$, so it too is reduced by the dielectric. Thus there is a smaller voltage V for the same charge Q; since $C = Q/V$, the capacitance C is greater.

The dielectric constant is generally defined to be $\kappa = E_0/E$, or the ratio of the electric field in a vacuum to that in the dielectric material, and is intimately related to the polarizability of the material.

Note:
Things Great and Small
The Submicroscopic Origin of Polarization
Polarization is a separation of charge within an atom or molecule. As has been noted, the planetary model of the atom pictures it as having a positive nucleus orbited by negative electrons, analogous to the planets orbiting the Sun. Although this model is not completely accurate, it is very helpful in explaining a vast range of phenomena and will be refined elsewhere, such as in Atomic Physics. The submicroscopic origin of polarization can be modeled as shown in [link].



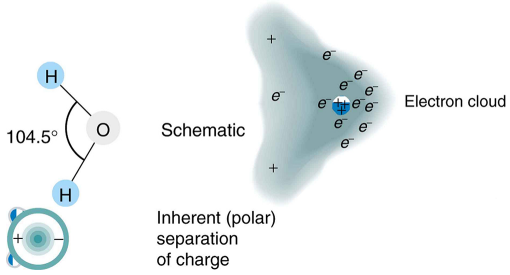
Artist's conception of a polarized atom. The orbits of

electrons around the nucleus are shifted slightly by the external charges (shown exaggerated). The resulting separation of charge within the atom means that it is polarized. Note that the unlike charge is now closer to the external charges, causing the polarization.

We will find in Atomic Physics that the orbits of electrons are more properly viewed as electron clouds with the density of the cloud related to the probability of finding an electron in that location (as opposed to the definite locations and paths of planets in their orbits around the Sun). This cloud is shifted by the Coulomb force so that the atom on average has a separation of charge. Although the atom remains neutral, it can now be the source of a Coulomb force, since a charge brought near the atom will be closer to one type of charge than the other.

Some molecules, such as those of water, have an inherent separation of charge and are thus called polar molecules. [link] illustrates the separation of charge in a water molecule, which has two hydrogen atoms and one oxygen atom (H_2O). The water molecule is not symmetric—the hydrogen atoms are repelled to one side, giving the molecule a boomerang shape. The electrons in a water molecule are more concentrated around the more highly charged oxygen nucleus than around the hydrogen nuclei. This makes the oxygen end of the molecule slightly negative and leaves the hydrogen ends slightly positive. The inherent separation of charge in polar molecules

makes it easier to align them with external fields and charges. Polar molecules therefore exhibit greater polarization effects and have greater dielectric constants. Those who study chemistry will find that the polar nature of water has many effects. For example, water molecules gather ions much more effectively because they have an electric field and a separation of charge to attract charges of both signs. Also, as brought out in the previous chapter, polar water provides a shield or screening of the electric fields in the highly charged molecules of interest in biological systems.



Artist’s conception of a water molecule. There is an inherent separation of charge, and so water is a polar molecule. Electrons in the molecule are attracted to the oxygen nucleus and leave an excess of positive charge near the two hydrogen nuclei. (Note that the schematic on the right is a rough illustration of the distribution of electrons in the water molecule. It does not show the actual numbers of protons and electrons involved in the structure.)

Note:
PhET Explorations: Capacitor Lab
Explore how a capacitor works! Change the size of the plates and add a dielectric to see the effect on capacitance. Change the voltage and see charges built up on the plates. Observe the electric field in the capacitor. Measure the voltage and the electric field.

Capacitor Lab

Section Summary

- A capacitor is a device used to store charge.
- The amount of charge Q a capacitor can store depends on two major factors—the voltage applied and the capacitor's physical characteristics, such as its size.
- The capacitance C is the amount of charge stored per volt, or
  Equation:

$$C = \frac{Q}{V}.$$

- The capacitance of a parallel plate capacitor is $C = \varepsilon_0 \frac{A}{d}$, when the plates are separated by air or free space. ε_0 is called the permittivity of free space.
- A parallel plate capacitor with a dielectric between its plates has a capacitance given by
  Equation:

$$C = \kappa\varepsilon_0 \frac{A}{d},$$

where κ is the dielectric constant of the material.

- The maximum electric field strength above which an insulating material begins to break down and conduct is called dielectric strength.

Conceptual Questions

Exercise:

Problem:

Does the capacitance of a device depend on the applied voltage? What about the charge stored in it?

Exercise:

Problem:

Use the characteristics of the Coulomb force to explain why capacitance should be proportional to the plate area of a capacitor. Similarly, explain why capacitance should be inversely proportional to the separation between plates.

Exercise:

Problem:

Give the reason why a dielectric material increases capacitance compared with what it would be with air between the plates of a capacitor. What is the independent reason that a dielectric material also allows a greater voltage to be applied to a capacitor? (The dielectric thus increases C and permits a greater V.)

Exercise:

Problem:

How does the polar character of water molecules help to explain water's relatively large dielectric constant? ([link])

Exercise:

Problem:

Sparks will occur between the plates of an air-filled capacitor at lower voltage when the air is humid than when dry. Explain why, considering the polar character of water molecules.

Exercise:

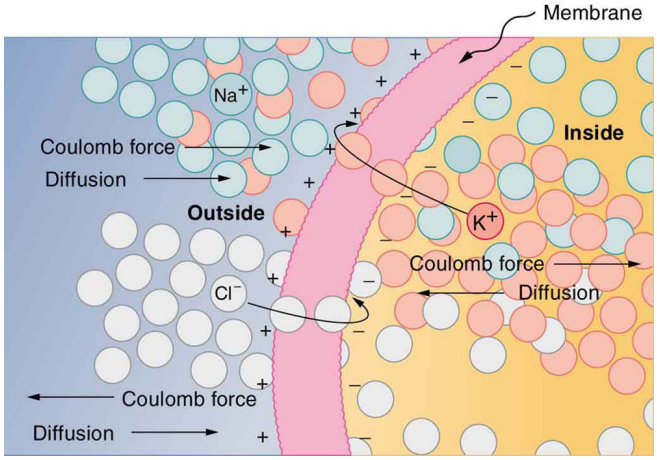
Problem:

Water has a large dielectric constant, but it is rarely used in capacitors. Explain why.

Exercise:

Problem:

Membranes in living cells, including those in humans, are characterized by a separation of charge across the membrane. Effectively, the membranes are thus charged capacitors with important functions related to the potential difference across the membrane. Is energy required to separate these charges in living membranes and, if so, is its source the metabolization of food energy or some other source?



The semipermeable membrane of a cell has different concentrations of ions inside and out. Diffusion moves the K^+ (potassium) and Cl^- (chloride) ions in the directions shown, until the Coulomb force halts further transfer. This results in a layer of positive charge on the outside, a layer of negative charge on the inside, and thus a voltage across the cell membrane. The membrane is normally impermeable to Na^+ (sodium ions).

Problems & Exercises

Exercise:

Problem:

What charge is stored in a $180\ \mu F$ capacitor when 120 V is applied to it?

Solution:

$21.6\ mC$

Exercise:

Problem:

Find the charge stored when 5.50 V is applied to an 8.00 pF capacitor.

Exercise:

Problem: What charge is stored in the capacitor in [link]?

Solution:

80.0 mC

Exercise:

Problem:

Calculate the voltage applied to a 2.00 μF capacitor when it holds 3.10 μC of charge.

Exercise:

Problem:

What voltage must be applied to an 8.00 nF capacitor to store 0.160 mC of charge?

Solution:

20.0 kV

Exercise:

Problem:

What capacitance is needed to store 3.00 μC of charge at a voltage of 120 V?

Exercise:

Problem:

What is the capacitance of a large Van de Graaff generator’s terminal, given that it stores 8.00 mC of charge at a voltage of 12.0 MV?

Solution:

667 pF

Exercise:

Problem:

Find the capacitance of a parallel plate capacitor having plates of area $5.00\ \mathrm{m}^2$ that are separated by 0.100 mm of Teflon.

Exercise:

Problem:

(a)What is the capacitance of a parallel plate capacitor having plates of area $1.50\ \mathrm{m}^2$ that are separated by 0.0200 mm of neoprene rubber? (b) What charge does it hold when 9.00 V is applied to it?

Solution:

(a) $4.4\ \mu\mathrm{F}$

(b) $4.0 \times 10^{-5}\ \mathrm{C}$

Exercise:

Problem: Integrated Concepts

A prankster applies 450 V to an $80.0\ \mu\mathrm{F}$ capacitor and then tosses it to an unsuspecting victim. The victim's finger is burned by the discharge of the capacitor through 0.200 g of flesh. What is the temperature increase of the flesh? Is it reasonable to assume no phase change?

Exercise:

Problem:Unreasonable Results

(a) A certain parallel plate capacitor has plates of area $4.00\ \mathrm{m}^2$, separated by 0.0100 mm of nylon, and stores 0.170 C of charge. What is the applied voltage? (b) What is unreasonable about this result? (c) Which assumptions are responsible or inconsistent?

Solution:

(a) 14.2 kV

(b) The voltage is unreasonably large, more than 100 times the breakdown voltage of nylon.

(c) The assumed charge is unreasonably large and cannot be stored in a capacitor of these dimensions.

Glossary

capacitor
: a device that stores electric charge

capacitance
: amount of charge stored per unit volt

dielectric
: an insulating material

dielectric strength
: the maximum electric field above which an insulating material begins to break down and conduct

parallel plate capacitor
: two identical conducting plates separated by a distance

polar molecule
: a molecule with inherent separation of charge

DC Circuits Containing Resistors and Capacitors

- Explain the importance of the time constant, τ , and calculate the time constant for a given resistance and capacitance.
- Explain why batteries in a flashlight gradually lose power and the light dims over time.
- Describe what happens to a graph of the voltage across a capacitor over time as it charges.
- Explain how a timing circuit works and list some applications.
- Calculate the necessary speed of a strobe flash needed to “stop” the movement of an object over a particular length.

When you use a flash camera, it takes a few seconds to charge the capacitor that powers the flash. The light flash discharges the capacitor in a tiny fraction of a second. Why does charging take longer than discharging? This question and a number of other phenomena that involve charging and discharging capacitors are discussed in this module.

RC Circuits

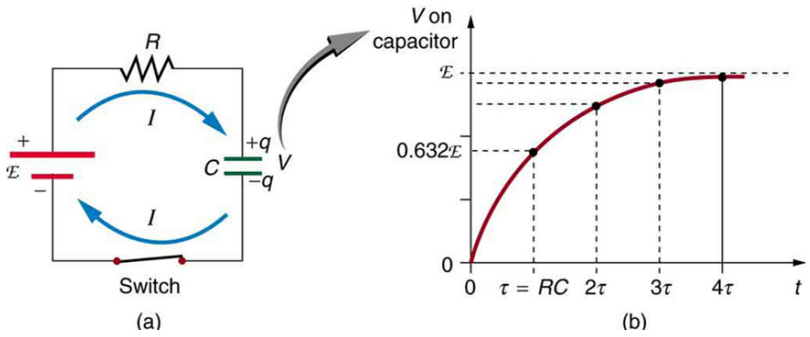
An **RC circuit** is one containing a **resistor** R and a **capacitor** C. The capacitor is an electrical component that stores electric charge.

[link] shows a simple RC circuit that employs a DC (direct current) voltage source. The capacitor is initially uncharged. As soon as the switch is closed, current flows to and from the initially uncharged capacitor. As charge increases on the capacitor plates, there is increasing opposition to the flow of charge by the repulsion of like charges on each plate.

In terms of voltage, this is because voltage across the capacitor is given by $V_c = Q/C$, where Q is the amount of charge stored on each plate and C is the **capacitance**. This voltage opposes the battery, growing from zero to the maximum emf when fully charged. The current thus decreases from its initial value of $I_0 = \frac{\text{emf}}{R}$ to zero as the voltage on the capacitor reaches the same value as the emf. When there is no current, there is no IR drop, and so the voltage on the capacitor must then equal the emf of the voltage source. This can also be explained with Kirchhoff’s second rule (the loop rule),

discussed in Kirchhoff's Rules, which says that the algebraic sum of changes in potential around any closed loop must be zero.

The initial current is $I_0 = \frac{\text{emf}}{R}$, because all of the IR drop is in the resistance. Therefore, the smaller the resistance, the faster a given capacitor will be charged. Note that the internal resistance of the voltage source is included in R, as are the resistances of the capacitor and the connecting wires. In the flash camera scenario above, when the batteries powering the camera begin to wear out, their internal resistance rises, reducing the current and lengthening the time it takes to get ready for the next flash.



(a) An RC circuit with an initially uncharged capacitor. Current flows in the direction shown (opposite of electron flow) as soon as the switch is closed. Mutual repulsion of like charges in the capacitor progressively slows the flow as the capacitor is charged, stopping the current when the capacitor is fully charged and $Q = C \cdot \text{emf}$. (b) A graph of voltage across the capacitor versus time, with the switch closing at time $t = 0$. (Note that in the two parts of the figure, the capital script E stands for emf, q stands for the charge stored on the capacitor, and τ is the RC time constant.)

Voltage on the capacitor is initially zero and rises rapidly at first, since the initial current is a maximum. [link](b) shows a graph of capacitor voltage versus time (t) starting when the switch is closed at $t = 0$. The voltage approaches emf asymptotically, since the closer it gets to emf the less current flows. The equation for voltage versus time when charging a capacitor C through a resistor R, derived using calculus, is
Equation:

$$V = \text{emf}(1 - e^{-t/\text{RC}}) \text{ (charging)},$$

where V is the voltage across the capacitor, emf is equal to the emf of the DC voltage source, and the exponential e = 2.718 … is the base of the natural logarithm. Note that the units of RC are seconds. We define
Equation:

$$\tau = \text{RC},$$

where τ (the Greek letter tau) is called the time constant for an RC circuit. As noted before, a small resistance R allows the capacitor to charge faster. This is reasonable, since a larger current flows through a smaller resistance. It is also reasonable that the smaller the capacitor C, the less time needed to charge it. Both factors are contained in $\tau = \text{RC}$.

More quantitatively, consider what happens when $t = \tau = \text{RC}$. Then the voltage on the capacitor is
Equation:

$$V = \text{emf}\left(1 - e^{-1}\right) = \text{emf}(1 - 0.368) = 0.632 \cdot \text{emf}.$$

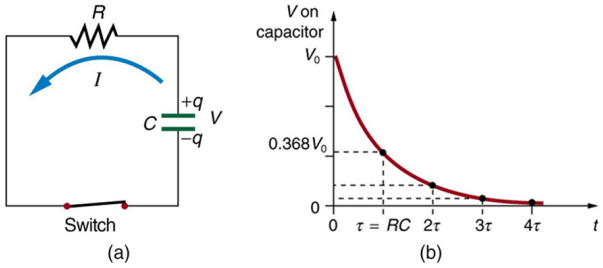
This means that in the time $\tau = \text{RC}$, the voltage rises to 0.632 of its final value. The voltage will rise 0.632 of the remainder in the next time τ. It is a characteristic of the exponential function that the final value is never reached, but 0.632 of the remainder to that value is achieved in every time, τ. In just a few multiples of the time constant τ, then, the final value is very nearly achieved, as the graph in [link](b) illustrates.

Discharging a Capacitor

Discharging a capacitor through a resistor proceeds in a similar fashion, as [link] illustrates. Initially, the current is $I_0 = \frac{V_0}{R}$, driven by the initial voltage V_0 on the capacitor. As the voltage decreases, the current and hence the rate of discharge decreases, implying another exponential formula for V. Using calculus, the voltage V on a capacitor C being discharged through a resistor R is found to be

Equation:

$$V = V_0\, e^{-t/\mathrm{RC}} (\text{discharging}).$$



(a) Closing the switch discharges the capacitor C through the resistor R. Mutual repulsion of like charges on each plate drives the current. (b) A graph of voltage across the capacitor versus time, with $V = V_0$ at $t = 0$. The voltage decreases exponentially, falling a fixed fraction of the way to zero in each subsequent time constant τ.

The graph in [link](b) is an example of this exponential decay. Again, the time constant is $\tau = \mathrm{RC}$. A small resistance R allows the capacitor to discharge in a small time, since the current is larger. Similarly, a small capacitance requires less time to discharge, since less charge is stored. In the first time interval $\tau = \mathrm{RC}$ after the switch is closed, the voltage falls to 0.368 of its initial value, since $V = V_0 \cdot e^{-1} = 0.368 V_0$.

During each successive time τ, the voltage falls to 0.368 of its preceding value. In a few multiples of τ, the voltage becomes very close to zero, as indicated by the graph in [link](b).

Now we can explain why the flash camera in our scenario takes so much longer to charge than discharge; the resistance while charging is significantly greater than while discharging. The internal resistance of the battery accounts for most of the resistance while charging. As the battery ages, the increasing internal resistance makes the charging process even slower. (You may have noticed this.)

The flash discharge is through a low-resistance ionized gas in the flash tube and proceeds very rapidly. Flash photographs, such as in [link], can capture a brief instant of a rapid motion because the flash can be less than a microsecond in duration. Such flashes can be made extremely intense.

During World War II, nighttime reconnaissance photographs were made from the air with a single flash illuminating more than a square kilometer of enemy territory. The brevity of the flash eliminated blurring due to the surveillance aircraft’s motion. Today, an important use of intense flash lamps is to pump energy into a laser. The short intense flash can rapidly energize a laser and allow it to reemit the energy in another form.



This stop-motion photograph of a rufous hummingbird (*Selasphorus rufus*) feeding on a flower was obtained with an extremely brief and intense flash of light powered by the discharge of a capacitor through a gas. (credit: Dean E. Biggins, U.S. Fish and Wildlife Service)

Example:
Integrated Concept Problem: Calculating Capacitor Size—Strobe Lights
High-speed flash photography was pioneered by Doc Edgerton in the 1930s, while he was a professor of electrical engineering at MIT. You might have seen examples of his work in the amazing shots of hummingbirds in motion, a drop of milk splattering on a table, or a bullet penetrating an apple (see [link]). To stop the motion and capture these pictures, one needs a high-intensity, very short pulsed flash, as mentioned earlier in this module.
Suppose one wished to capture the picture of a bullet (moving at $5.0 \times 10^2\ \mathrm{m/s}$) that was passing through an apple. The duration of the flash is related to the RC time constant, τ. What size capacitor would one

need in the RC circuit to succeed, if the resistance of the flash tube was 10.0 Ω? Assume the apple is a sphere with a diameter of 8.0×10^{-2} m.

Strategy

We begin by identifying the physical principles involved. This example deals with the strobe light, as discussed above. [link] shows the circuit for this probe. The characteristic time τ of the strobe is given as $\tau = \mathrm{RC}$.

Solution

We wish to find C, but we don't know τ. We want the flash to be on only while the bullet traverses the apple. So we need to use the kinematic equations that describe the relationship between distance x, velocity v, and time t:

Equation:

$$x = \mathrm{vt} \text{ or } t = \frac{x}{v}.$$

The bullet's velocity is given as 5.0×10^2 m/s, and the distance x is 8.0×10^{-2} m. The traverse time, then, is

Equation:

$$t = \frac{x}{v} = \frac{8.0 \times 10^{-2} \text{ m}}{5.0 \times 10^2 \text{ m/s}} = 1.6 \times 10^{-4} \text{ s}.$$

We set this value for the crossing time t equal to τ. Therefore,

Equation:

$$C = \frac{t}{R} = \frac{1.6 \times 10^{-4} \text{ s}}{10.0 \ \Omega} = 16 \ \mu\text{F}.$$

(Note: Capacitance C is typically measured in farads, F, defined as Coulombs per volt. From the equation, we see that C can also be stated in units of seconds per ohm.)

Discussion

The flash interval of 160 μs (the traverse time of the bullet) is relatively easy to obtain today. Strobe lights have opened up new worlds from science to entertainment. The information from the picture of the apple and bullet was used in the Warren Commission Report on the assassination of

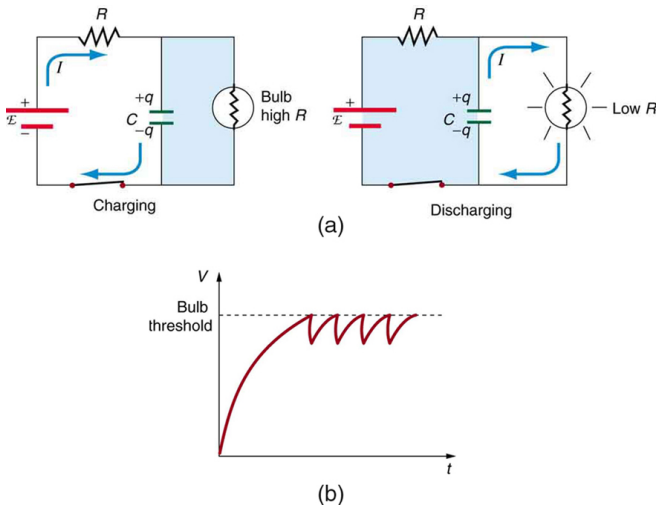
President John F. Kennedy in 1963 to confirm that only one bullet was fired.

RC Circuits for Timing

RC circuits are commonly used for timing purposes. A mundane example of this is found in the ubiquitous intermittent wiper systems of modern cars. The time between wipes is varied by adjusting the resistance in an RC circuit. Another example of an RC circuit is found in novelty jewelry, Halloween costumes, and various toys that have battery-powered flashing lights. (See [link] for a timing circuit.)

A more crucial use of RC circuits for timing purposes is in the artificial pacemaker, used to control heart rate. The heart rate is normally controlled by electrical signals generated by the sino-atrial (SA) node, which is on the wall of the right atrium chamber. This causes the muscles to contract and pump blood. Sometimes the heart rhythm is abnormal and the heartbeat is too high or too low.

The artificial pacemaker is inserted near the heart to provide electrical signals to the heart when needed with the appropriate time constant. Pacemakers have sensors that detect body motion and breathing to increase the heart rate during exercise to meet the body’s increased needs for blood and oxygen.



(a) The lamp in this RC circuit ordinarily has a very high resistance, so that the battery charges the capacitor as if the lamp were not there. When the voltage reaches a threshold value, a current flows through the lamp that dramatically reduces its resistance, and the capacitor discharges through the lamp as if the battery and charging resistor were not there. Once discharged, the process starts again, with the flash period determined by the RC constant τ. (b) A graph of voltage versus time for this circuit.

Example:
Calculating Time: *RC* Circuit in a Heart Defibrillator
A heart defibrillator is used to resuscitate an accident victim by discharging a capacitor through the trunk of her body. A simplified version of the

circuit is seen in [link]. (a) What is the time constant if an 8.00-μF capacitor is used and the path resistance through her body is $1.00 \times 10^3\ \Omega$? (b) If the initial voltage is 10.0 kV, how long does it take to decline to 5.00×10^2 V?

Strategy

Since the resistance and capacitance are given, it is straightforward to multiply them to give the time constant asked for in part (a). To find the time for the voltage to decline to 5.00×10^2 V, we repeatedly multiply the initial voltage by 0.368 until a voltage less than or equal to 5.00×10^2 V is obtained. Each multiplication corresponds to a time of τ seconds.

Solution for (a)

The time constant τ is given by the equation $\tau = \mathrm{RC}$. Entering the given values for resistance and capacitance (and remembering that units for a farad can be expressed as s/Ω) gives

Equation:

$$\tau = \mathrm{RC} = (1.00 \times 10^3\ \Omega)(8.00\ \mu F) = 8.00\ \mathrm{ms}.$$

Solution for (b)

In the first 8.00 ms, the voltage (10.0 kV) declines to 0.368 of its initial value. That is:

Equation:

$$V = 0.368 V_0 = 3.680 \times 10^3\ \mathrm{V\ at}\ t = 8.00\ \mathrm{ms}.$$

(Notice that we carry an extra digit for each intermediate calculation.) After another 8.00 ms, we multiply by 0.368 again, and the voltage is

Equation:

$$\begin{aligned} V\prime &= 0.368V \\ &= (0.368)\left(3.680 \times 10^3\ \mathrm{V}\right) \\ &= 1.354 \times 10^3\ \mathrm{V\ at}\ t = 16.0\ \mathrm{ms}. \end{aligned}$$

Similarly, after another 8.00 ms, the voltage is

Equation:

$$\begin{aligned} V'' &= 0.368V' = (0.368)(1.354 \times 10^3 \text{ V}) \\ &= 498 \text{ V at } t = 24.0 \text{ ms}. \end{aligned}$$

Discussion
So after only 24.0 ms, the voltage is down to 498 V, or 4.98% of its original value.Such brief times are useful in heart defibrillation, because the brief but intense current causes a brief but effective contraction of the heart. The actual circuit in a heart defibrillator is slightly more complex than the one in [link], to compensate for magnetic and AC effects that will be covered in Magnetism.

Exercise:
Check Your Understanding

Problem: When is the potential difference across a capacitor an emf?

Solution:

Only when the current being drawn from or put into the capacitor is zero. Capacitors, like batteries, have internal resistance, so their output voltage is not an emf unless current is zero. This is difficult to measure in practice so we refer to a capacitor's voltage rather than its emf. But the source of potential difference in a capacitor is fundamental and it is an emf.

Note:
PhET Explorations: Circuit Construction Kit (DC only)
An electronics kit in your computer! Build circuits with resistors, light bulbs, batteries, and switches. Take measurements with the realistic ammeter and voltmeter. View the circuit as a schematic diagram, or switch to a life-like view.
https://archive.cnx.org/specials/f23ce496-c9d1-11e5-bdc8-bb04dc1eecb6/circuit-construction-kit-dc-only/#sim-cck

Section Summary

- An RC circuit is one that has both a resistor and a capacitor.
- The time constant τ for an RC circuit is $\tau = \mathrm{RC}$.
- When an initially uncharged ($V_0 = 0$ at $t = 0$) capacitor in series with a resistor is charged by a DC voltage source, the voltage rises, asymptotically approaching the emf of the voltage source; as a function of time,
  Equation:

$$V = \mathrm{emf}(1 - e^{-t/\mathrm{RC}})(\mathrm{charging}).$$

- Within the span of each time constant τ, the voltage rises by 0.632 of the remaining value, approaching the final voltage asymptotically.
- If a capacitor with an initial voltage V_0 is discharged through a resistor starting at $t = 0$, then its voltage decreases exponentially as given by
  Equation:

$$V = V_0 e^{-t/\mathrm{RC}}(\mathrm{discharging}).$$

- In each time constant τ, the voltage falls by 0.368 of its remaining initial value, approaching zero asymptotically.

Conceptual questions

Exercise:

Problem:

Regarding the units involved in the relationship $\tau = \mathrm{RC}$, verify that the units of resistance times capacitance are time, that is, $\Omega \cdot \mathrm{F} = \mathrm{s}$.

Exercise:

Problem:

The RC time constant in heart defibrillation is crucial to limiting the time the current flows. If the capacitance in the defibrillation unit is fixed, how would you manipulate resistance in the circuit to adjust the RC constant τ? Would an adjustment of the applied voltage also be needed to ensure that the current delivered has an appropriate value?

Exercise:

Problem:

When making an ECG measurement, it is important to measure voltage variations over small time intervals. The time is limited by the RC constant of the circuit—it is not possible to measure time variations shorter than RC. How would you manipulate R and C in the circuit to allow the necessary measurements?

Exercise:

Problem:

Draw two graphs of charge versus time on a capacitor. Draw one for charging an initially uncharged capacitor in series with a resistor, as in the circuit in [link], starting from $\mathrm{t} = 0$. Draw the other for discharging a capacitor through a resistor, as in the circuit in [link], starting at $\mathrm{t} = 0$, with an initial charge Q_0. Show at least two intervals of τ.

Exercise:

Problem:

When charging a capacitor, as discussed in conjunction with [link], how long does it take for the voltage on the capacitor to reach emf? Is this a problem?

Exercise:

Problem:

When discharging a capacitor, as discussed in conjunction with [link], how long does it take for the voltage on the capacitor to reach zero? Is this a problem?

Exercise:

Problem:

Referring to [link], draw a graph of potential difference across the resistor versus time, showing at least two intervals of τ. Also draw a graph of current versus time for this situation.

Exercise:

Problem:

A long, inexpensive extension cord is connected from inside the house to a refrigerator outside. The refrigerator doesn't run as it should. What might be the problem?

Exercise:

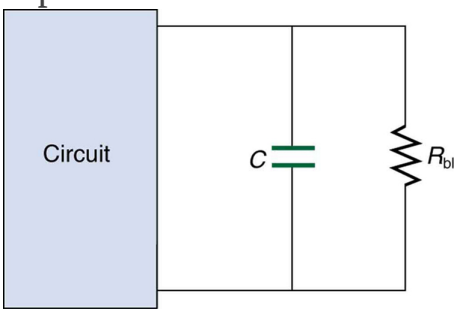
Problem:

In [link], does the graph indicate the time constant is shorter for discharging than for charging? Would you expect ionized gas to have low resistance? How would you adjust R to get a longer time between flashes? Would adjusting R affect the discharge time?

Exercise:

Problem:

An electronic apparatus may have large capacitors at high voltage in the power supply section, presenting a shock hazard even when the apparatus is switched off. A “bleeder resistor” is therefore placed across such a capacitor, as shown schematically in [link], to bleed the charge from it after the apparatus is off. Why must the bleeder resistance be much greater than the effective resistance of the rest of the circuit? How does this affect the time constant for discharging the capacitor?



A bleeder resistor R_{bl} discharges the capacitor in this electronic device once it is switched off.

Problem Exercises

Exercise:

Problem:

The timing device in an automobile’s intermittent wiper system is based on an RC time constant and utilizes a $0.500\text{-}\mu\text{F}$ capacitor and a variable resistor. Over what range must R be made to vary to achieve time constants from 2.00 to 15.0 s?

Solution:

range 4.00 to 30.0 MΩ

Exercise:

Problem:

A heart pacemaker fires 72 times a minute, each time a 25.0-nF capacitor is charged (by a battery in series with a resistor) to 0.632 of its full voltage. What is the value of the resistance?

Exercise:

Problem:

The duration of a photographic flash is related to an RC time constant, which is 0.100 μs for a certain camera. (a) If the resistance of the flash lamp is 0.0400 Ω during discharge, what is the size of the capacitor supplying its energy? (b) What is the time constant for charging the capacitor, if the charging resistance is 800 kΩ?

Solution:

(a) 2.50 μF

(b) 2.00 s

Exercise:

Problem:

A 2.00- and a 7.50-μF capacitor can be connected in series or parallel, as can a 25.0- and a 100-kΩ resistor. Calculate the four RC time constants possible from connecting the resulting capacitance and resistance in series.

Exercise:

Problem:

After two time constants, what percentage of the final voltage, emf, is on an initially uncharged capacitor C, charged through a resistance R?

Solution:

86.5%

Exercise:

Problem:

A $500\text{-}\Omega$ resistor, an uncharged $1.50\text{-}\mu\text{F}$ capacitor, and a 6.16-V emf are connected in series. (a) What is the initial current? (b) What is the RC time constant? (c) What is the current after one time constant? (d) What is the voltage on the capacitor after one time constant?

Exercise:

Problem:

A heart defibrillator being used on a patient has an RC time constant of 10.0 ms due to the resistance of the patient and the capacitance of the defibrillator. (a) If the defibrillator has an $8.00\text{-}\mu\text{F}$ capacitance, what is the resistance of the path through the patient? (You may neglect the capacitance of the patient and the resistance of the defibrillator.) (b) If the initial voltage is 12.0 kV, how long does it take to decline to 6.00×10^2 V?

Solution:

(a) $1.25\ \text{k}\Omega$

(b) 30.0 ms

Exercise:

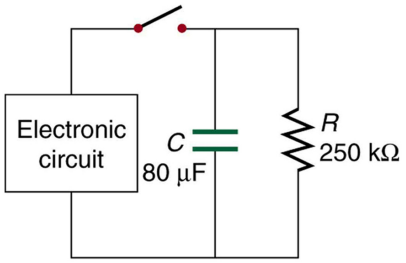
Problem:

An ECG monitor must have an RC time constant less than 1.00×10^2 μs to be able to measure variations in voltage over small time intervals. (a) If the resistance of the circuit (due mostly to that of the patient's chest) is 1.00 kΩ, what is the maximum capacitance of the circuit? (b) Would it be difficult in practice to limit the capacitance to less than the value found in (a)?

Exercise:

Problem:

[link] shows how a bleeder resistor is used to discharge a capacitor after an electronic device is shut off, allowing a person to work on the electronics with less risk of shock. (a) What is the time constant? (b) How long will it take to reduce the voltage on the capacitor to 0.250% (5% of 5%) of its full value once discharge begins? (c) If the capacitor is charged to a voltage V_0 through a 100-Ω resistance, calculate the time it takes to rise to $0.865V_0$ (This is about two time constants.)



Solution:

(a) 20.0 s

(b) 120 s

(c) 16.0 ms

Exercise:

Problem:

Using the exact exponential treatment, find how much time is required to discharge a $250\text{-}\mu\text{F}$ capacitor through a $500\text{-}\Omega$ resistor down to 1.00% of its original voltage.

Exercise:

Problem:

Using the exact exponential treatment, find how much time is required to charge an initially uncharged 100-pF capacitor through a $75.0\text{-}\text{M}\Omega$ resistor to 90.0% of its final voltage.

Solution:

$1.73 \times 10^{-2}\ \text{s}$

Exercise:

Problem: **Integrated Concepts**

If you wish to take a picture of a bullet traveling at 500 m/s, then a very brief flash of light produced by an RC discharge through a flash tube can limit blurring. Assuming 1.00 mm of motion during one RC constant is acceptable, and given that the flash is driven by a $600\text{-}\mu\text{F}$ capacitor, what is the resistance in the flash tube?

Solution:

$3.33\times10^{-3}\ \Omega$

Exercise:

Problem: **Integrated Concepts**

A flashing lamp in a Christmas earring is based on an RC discharge of a capacitor through its resistance. The effective duration of the flash is 0.250 s, during which it produces an average 0.500 W from an average

3.00 V. (a) What energy does it dissipate? (b) How much charge moves through the lamp? (c) Find the capacitance. (d) What is the resistance of the lamp?

Exercise:

Problem: Integrated Concepts

A 160-μF capacitor charged to 450 V is discharged through a 31.2-kΩ resistor. (a) Find the time constant. (b) Calculate the temperature increase of the resistor, given that its mass is 2.50 g and its specific heat is $1.67\,\frac{\text{kJ}}{\text{kg}\cdot{}^{\circ}\text{C}}$, noting that most of the thermal energy is retained in the short time of the discharge. (c) Calculate the new resistance, assuming it is pure carbon. (d) Does this change in resistance seem significant?

Solution:

(a) 4.99 s

(b) 3.87°C

(c) 31.1 kΩ

(d) No

Exercise:

Problem: Unreasonable Results

(a) Calculate the capacitance needed to get an RC time constant of 1.00×10^3 s with a 0.100-Ω resistor. (b) What is unreasonable about this result? (c) Which assumptions are responsible?

Exercise:

Problem: Construct Your Own Problem

Consider a camera’s flash unit. Construct a problem in which you calculate the size of the capacitor that stores energy for the flash lamp. Among the things to be considered are the voltage applied to the capacitor, the energy needed in the flash and the associated charge needed on the capacitor, the resistance of the flash lamp during discharge, and the desired RC time constant.

Exercise:

Problem: Construct Your Own Problem

Consider a rechargeable lithium cell that is to be used to power a camcorder. Construct a problem in which you calculate the internal resistance of the cell during normal operation. Also, calculate the minimum voltage output of a battery charger to be used to recharge your lithium cell. Among the things to be considered are the emf and useful terminal voltage of a lithium cell and the current it should be able to supply to a camcorder.

Glossary

RC circuit
: a circuit that contains both a resistor and a capacitor

capacitor
: an electrical component used to store energy by separating electric charge on two opposing plates

capacitance
: the maximum amount of electric potential energy that can be stored (or separated) for a given electric potential

Capacitors in Series and Parallel

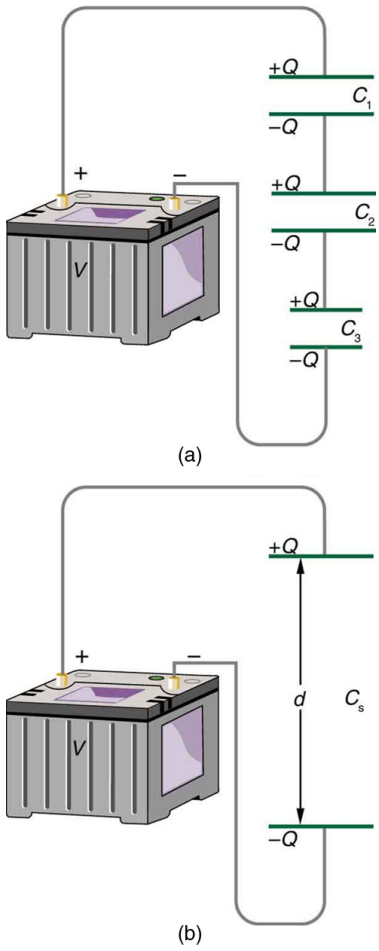
- Derive expressions for total capacitance in series and in parallel.
- Identify series and parallel parts in the combination of connection of capacitors.
- Calculate the effective capacitance in series and parallel given individual capacitances.

Several capacitors may be connected together in a variety of applications. Multiple connections of capacitors act like a single equivalent capacitor. The total capacitance of this equivalent single capacitor depends both on the individual capacitors and how they are connected. There are two simple and common types of connections, called *series* and *parallel*, for which we can easily calculate the total capacitance. Certain more complicated connections can also be related to combinations of series and parallel.

Capacitance in Series

[link](a) shows a series connection of three capacitors with a voltage applied. As for any capacitor, the capacitance of the combination is related to charge and voltage by $C = \frac{Q}{V}$.

Note in [link] that opposite charges of magnitude Q flow to either side of the originally uncharged combination of capacitors when the voltage V is applied. Conservation of charge requires that equal-magnitude charges be created on the plates of the individual capacitors, since charge is only being separated in these originally neutral devices. The end result is that the combination resembles a single capacitor with an effective plate separation greater than that of the individual capacitors alone. (See [link](b).) Larger plate separation means smaller capacitance. It is a general feature of series connections of capacitors that the total capacitance is less than any of the individual capacitances.



(a) Capacitors connected in series. The magnitude of the charge on each plate is Q. (b) An equivalent capacitor has a larger plate separation d. Series connections produce a total capacitance that is less than that of

any of the individual capacitors.

We can find an expression for the total capacitance by considering the voltage across the individual capacitors shown in [link]. Solving $C = \frac{Q}{V}$ for V gives $V = \frac{Q}{C}$. The voltages across the individual capacitors are thus $V_1 = \frac{Q}{C_1}$, $V_2 = \frac{Q}{C_2}$, and $V_3 = \frac{Q}{C_3}$. The total voltage is the sum of the individual voltages:

Equation:

$$V = V_1 + V_2 + V_3.$$

Now, calling the total capacitance C_{S} for series capacitance, consider that

Equation:

$$V = \frac{Q}{C_{\text{S}}} = V_1 + V_2 + V_3.$$

Entering the expressions for V_1, V_2, and V_3, we get

Equation:

$$\frac{Q}{C_{\text{S}}} = \frac{Q}{C_1} + \frac{Q}{C_2} + \frac{Q}{C_3}.$$

Canceling the Qs, we obtain the equation for the total capacitance in series C_{S} to be

Equation:

$$\frac{1}{C_{\text{S}}} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + ...,$$

where "..." indicates that the expression is valid for any number of capacitors connected in series. An expression of this form always results in a total capacitance C_{S} that is less than any of the individual capacitances C_1, C_2, ..., as the next example illustrates.

Note:
Total Capacitance in Series, C_{s}
Total capacitance in series: $\frac{1}{C_{\text{S}}} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \ldots$

Example:
What Is the Series Capacitance?
Find the total capacitance for three capacitors connected in series, given their individual capacitances are 1.000, 5.000, and 8.000 μF.
Strategy
With the given information, the total capacitance can be found using the equation for capacitance in series.
Solution
Entering the given capacitances into the expression for $\frac{1}{C_{\text{S}}}$ gives $\frac{1}{C_S} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3}$.
Equation:

$$\frac{1}{C_{\text{S}}} = \frac{1}{1.000\ \mu\text{F}} + \frac{1}{5.000\ \mu\text{F}} + \frac{1}{8.000\ \mu\text{F}} = \frac{1.325}{\mu\text{F}}$$

Inverting to find C_{S} yields $C_{\text{S}} = \frac{\mu\text{F}}{1.325} = 0.755\ \mu\text{F}$.
Discussion
The total series capacitance C_{s} is less than the smallest individual capacitance, as promised. In series connections of capacitors, the sum is less than the parts. In fact, it is less than any individual. Note that it is sometimes possible, and more convenient, to solve an equation like the above by finding the least common denominator, which in this case (showing only whole-number calculations) is 40. Thus,

Equation:

$$\frac{1}{C_{\text{S}}} = \frac{40}{40\ \mu\text{F}} + \frac{8}{40\ \mu\text{F}} + \frac{5}{40\ \mu\text{F}} = \frac{53}{40\ \mu\text{F}},$$

so that

Equation:

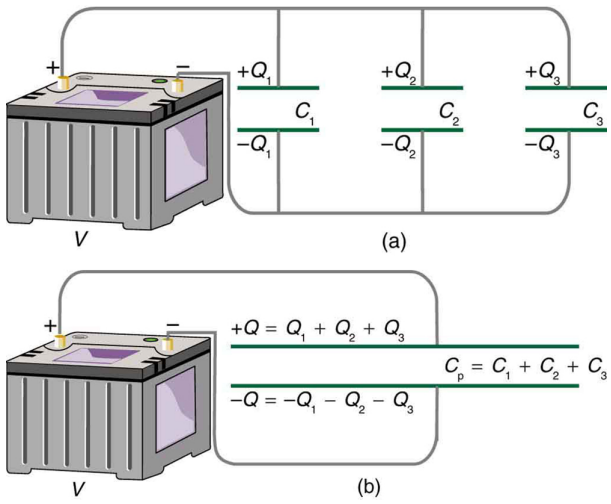
$$C_{\text{S}} = \frac{40\ \mu\text{F}}{53} = 0.755\ \mu\text{F}.$$

Capacitors in Parallel

[link](a) shows a parallel connection of three capacitors with a voltage applied. Here the total capacitance is easier to find than in the series case. To find the equivalent total capacitance C_{p}, we first note that the voltage across each capacitor is V, the same as that of the source, since they are connected directly to it through a conductor. (Conductors are equipotentials, and so the voltage across the capacitors is the same as that across the voltage source.) Thus the capacitors have the same charges on them as they would have if connected individually to the voltage source. The total charge Q is the sum of the individual charges:

Equation:

$$Q = Q_1 + Q_2 + Q_3.$$



(a) Capacitors in parallel. Each is connected directly to the voltage source just as if it were all alone, and so the total capacitance in parallel is just the sum of the individual capacitances. (b) The equivalent capacitor has a larger plate area and can therefore hold more charge than the individual capacitors.

Using the relationship $Q = \text{CV}$, we see that the total charge is $Q = C_\text{p}V$, and the individual charges are $Q_1 = C_1V$, $Q_2 = C_2V$, and $Q_3 = C_3V$. Entering these into the previous equation gives

Equation:

$$C_\text{p}V = C_1V + C_2V + C_3V.$$

Canceling V from the equation, we obtain the equation for the total capacitance in parallel C_p:

Equation:

$$C_{\text{p}} = C_1 + C_2 + C_3 + ...$$

Total capacitance in parallel is simply the sum of the individual capacitances. (Again the "..." indicates the expression is valid for any number of capacitors connected in parallel.) So, for example, if the capacitors in the example above were connected in parallel, their capacitance would be

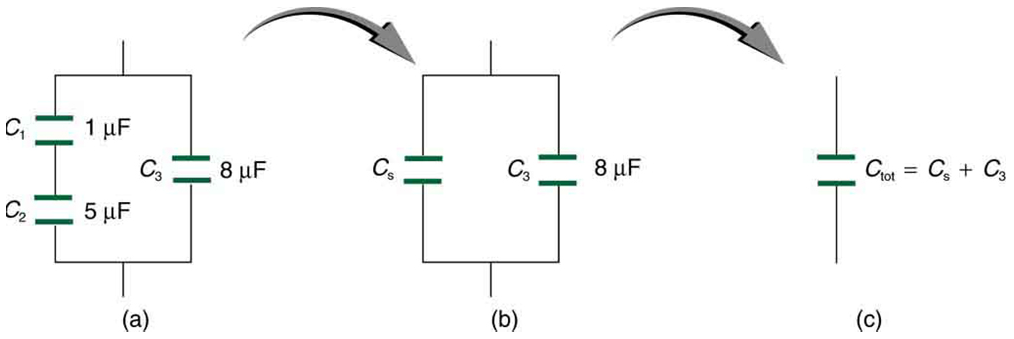
Equation:

$$C_{\text{p}} = 1.000\ \mu\text{F} + 5.000\ \mu\text{F} + 8.000\ \mu\text{F} = 14.000\ \mu\text{F}.$$

The equivalent capacitor for a parallel connection has an effectively larger plate area and, thus, a larger capacitance, as illustrated in [link](b).

Note:
Total Capacitance in Parallel, C_{p}
Total capacitance in parallel $C_{\text{p}} = C_1 + C_2 + C_3 + ...$

More complicated connections of capacitors can sometimes be combinations of series and parallel. (See [link].) To find the total capacitance of such combinations, we identify series and parallel parts, compute their capacitances, and then find the total.



(a) This circuit contains both series and parallel connections of capacitors. See [link] for the calculation of the overall capacitance of the circuit. (b) C_1 and C_2 are in series; their equivalent capacitance C_S is less than either of them. (c) Note that C_S is in parallel with C_3. The total capacitance is, thus, the sum of C_S and C_3.

Example:
A Mixture of Series and Parallel Capacitance
Find the total capacitance of the combination of capacitors shown in [link]. Assume the capacitances in [link] are known to three decimal places ($C_1 = 1.000\ \mu\text{F}$, $C_2 = 5.000\ \mu\text{F}$, and $C_3 = 8.000\ \mu\text{F}$), and round your answer to three decimal places.
Strategy
To find the total capacitance, we first identify which capacitors are in series and which are in parallel. Capacitors C_1 and C_2 are in series. Their combination, labeled C_S in the figure, is in parallel with C_3.
Solution
Since C_1 and C_2 are in series, their total capacitance is given by $\frac{1}{C_S} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3}$. Entering their values into the equation gives
Equation:

$$\frac{1}{C_S} = \frac{1}{C_1} + \frac{1}{C_2} = \frac{1}{1.000\ \mu\text{F}} + \frac{1}{5.000\ \mu\text{F}} = \frac{1.200}{\mu\text{F}}.$$

Inverting gives
Equation:

$$C_S = 0.833\ \mu\text{F}.$$

This equivalent series capacitance is in parallel with the third capacitor; thus, the total is the sum
Equation:

$$\begin{aligned} C_{\text{tot}} &= C_{\text{S}} + C_{\text{S}} \\ &= 0.833\ \mu\text{F} + 8.000\ \mu\text{F} \\ &= 8.833\ \mu\text{F}. \end{aligned}$$

Discussion
This technique of analyzing the combinations of capacitors piece by piece until a total is obtained can be applied to larger combinations of capacitors.

Section Summary

- Total capacitance in series $\frac{1}{C_{\text{S}}} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \ldots$
- Total capacitance in parallel $C_{\text{p}} = C_1 + C_2 + C_3 + \ldots$
- If a circuit contains a combination of capacitors in series and parallel, identify series and parallel parts, compute their capacitances, and then find the total.

Conceptual Questions

Exercise:

Problem:

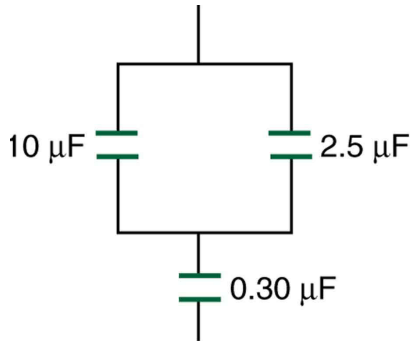
If you wish to store a large amount of energy in a capacitor bank, would you connect capacitors in series or parallel? Explain.

Problems & Exercises

Exercise:

Problem:

Find the total capacitance of the combination of capacitors in [link].



A combination of series and parallel connections of capacitors.

Solution:

$0.293\ \mu F$

Exercise:

Problem:

Suppose you want a capacitor bank with a total capacitance of 0.750 F and you possess numerous 1.50 mF capacitors. What is the smallest number you could hook together to achieve your goal, and how would you connect them?

Exercise:

Problem:

What total capacitances can you make by connecting a $5.00\ \mu F$ and an $8.00\ \mu F$ capacitor together?

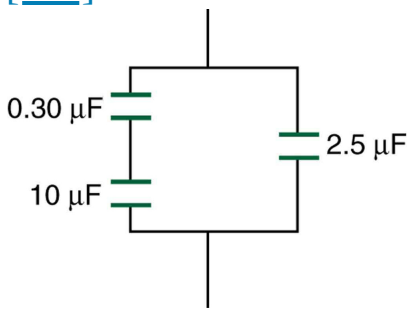
Solution:

$3.08\ \mu F$ in series combination, $13.0\ \mu F$ in parallel combination

Exercise:

Problem:

Find the total capacitance of the combination of capacitors shown in [link].



A combination of series and parallel connections of capacitors.

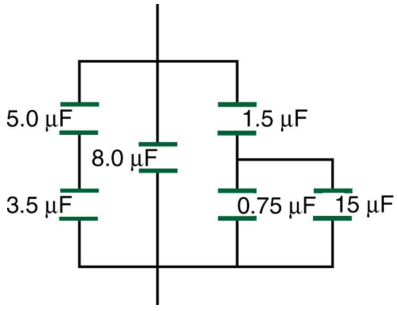
Solution:

2.79 μF

Exercise:

Problem:

Find the total capacitance of the combination of capacitors shown in [link].



A combination of series and parallel

connections of capacitors.

Exercise:

Problem: Unreasonable Results

(a) An $8.00\ \mu\text{F}$ capacitor is connected in parallel to another capacitor, producing a total capacitance of $5.00\ \mu\text{F}$. What is the capacitance of the second capacitor? (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Solution:

(a) $-3.00\ \mu\text{F}$

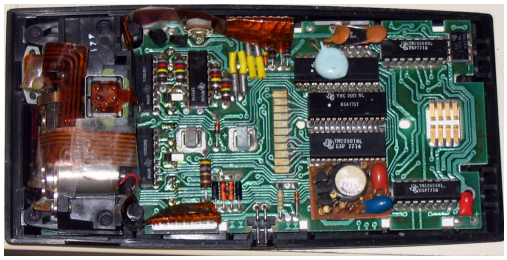
(b) You cannot have a negative value of capacitance.

(c) The assumption that the capacitors were hooked up in parallel, rather than in series, was incorrect. A parallel connection always produces a greater capacitance, while here a smaller capacitance was assumed. This could happen only if the capacitors are connected in series.

Energy Stored in Capacitors

- List some uses of capacitors.
- Express in equation form the energy stored in a capacitor.
- Explain the function of a defibrillator.

Most of us have seen dramatizations in which medical personnel use a **defibrillator** to pass an electric current through a patient's heart to get it to beat normally. (Review [link].) Often realistic in detail, the person applying the shock directs another person to "make it 400 joules this time." The energy delivered by the defibrillator is stored in a capacitor and can be adjusted to fit the situation. SI units of joules are often employed. Less dramatic is the use of capacitors in microelectronics, such as certain handheld calculators, to supply energy when batteries are charged. (See [link].) Capacitors are also used to supply energy for flash lamps on cameras.



Energy stored in the large capacitor is used to preserve the memory of an electronic calculator when its batteries are charged. (credit: Kucharek, Wikimedia Commons)

Energy stored in a capacitor is electrical potential energy, and it is thus related to the charge Q and voltage V on the capacitor. We must be careful when applying the equation for electrical potential energy $\Delta \mathrm{PE} = q\Delta V$ to a capacitor. Remember that $\Delta \mathrm{PE}$ is the potential energy of a charge q going through a voltage ΔV. But the capacitor starts with zero voltage and gradually comes up to its full voltage as it is charged. The first charge placed on a capacitor experiences a change in voltage $\Delta V = 0$, since the capacitor has zero voltage when uncharged. The final charge placed on a capacitor experiences $\Delta V = V$, since the capacitor now has its full voltage V on it. The average voltage on the capacitor during the charging process is $V/2$, and so the average voltage experienced by the full charge q is $V/2$. Thus the energy stored in a capacitor, E_{cap}, is

Equation:

$$E_{\mathrm{cap}} = \frac{QV}{2},$$

where Q is the charge on a capacitor with a voltage V applied. (Note that the energy is not QV, but $\mathrm{QV}/2$.) Charge and voltage are related to the capacitance C of a capacitor by $Q = \mathrm{CV}$, and so the expression for E_{cap} can be algebraically manipulated into three equivalent expressions:

Equation:

$$E_{\mathrm{cap}} = \frac{QV}{2} = \frac{CV^2}{2} = \frac{Q^2}{2C},$$

where Q is the charge and V the voltage on a capacitor C. The energy is in joules for a charge in coulombs, voltage in volts, and capacitance in farads.

Note:
Energy Stored in Capacitors
The energy stored in a capacitor can be expressed in three ways:
Equation:

$$E_{\text{cap}} = \frac{QV}{2} = \frac{CV^2}{2} = \frac{Q^2}{2C},$$

where Q is the charge, V is the voltage, and C is the capacitance of the capacitor. The energy is in joules for a charge in coulombs, voltage in volts, and capacitance in farads.

In a defibrillator, the delivery of a large charge in a short burst to a set of paddles across a person's chest can be a lifesaver. The person's heart attack might have arisen from the onset of fast, irregular beating of the heart—cardiac or ventricular fibrillation. The application of a large shock of electrical energy can terminate the arrhythmia and allow the body's pacemaker to resume normal patterns. Today it is common for ambulances to carry a defibrillator, which also uses an electrocardiogram to analyze the patient's heartbeat pattern. Automated external defibrillators (AED) are found in many public places ([link]). These are designed to be used by lay persons. The device automatically diagnoses the patient's heart condition and then applies the shock with appropriate energy and waveform. CPR is recommended in many cases before use of an AED.



Automated external defibrillators are found in many public places. These portable units provide verbal instructions for use in the important first few minutes for a person suffering a cardiac attack. (credit: Owain Davies, Wikimedia Commons)

Example:
Capacitance in a Heart Defibrillator
A heart defibrillator delivers 4.00×10^2 J of energy by discharging a capacitor initially at 1.00×10^4 V. What is its capacitance?
Strategy
We are given E_{cap} and V, and we are asked to find the capacitance C. Of the three expressions in the equation for E_{cap}, the most convenient relationship is
Equation:

$$E_{\text{cap}} = \frac{CV^2}{2}.$$

Solution
Solving this expression for C and entering the given values yields
Equation:

$$\begin{aligned} C &= \frac{2E_{\text{cap}}}{V^2} = \frac{2(4.00 \times 10^2 \text{ J})}{(1.00 \times 10^4 \text{ V})^2} = 8.00 \times 10^{-6} \text{ F} \\ &= 8.00 \text{ μF}. \end{aligned}$$

Discussion
This is a fairly large, but manageable, capacitance at 1.00×10^4 V.

Section Summary

- Capacitors are used in a variety of devices, including defibrillators, microelectronics such as calculators, and flash lamps, to supply energy.
- The energy stored in a capacitor can be expressed in three ways:
  Equation:

$$E_{\text{cap}} = \frac{\text{QV}}{2} = \frac{CV^2}{2} = \frac{Q^2}{2C},$$

  where Q is the charge, V is the voltage, and C is the capacitance of the capacitor. The energy is in joules when the charge is in coulombs, voltage is in volts, and capacitance is in farads.

Conceptual Questions

Exercise:

Problem:

How does the energy contained in a charged capacitor change when a dielectric is inserted, assuming the capacitor is isolated and its charge is constant? Does this imply that work was done?

Exercise:

Problem:

What happens to the energy stored in a capacitor connected to a battery when a dielectric is inserted? Was work done in the process?

Problems & Exercises

Exercise:

Problem:

(a) What is the energy stored in the $10.0\ \mu\text{F}$ capacitor of a heart defibrillator charged to $9.00 \times 10^3\ \text{V}$? (b) Find the amount of stored charge.

Solution:

(a) $405\ \text{J}$

(b) $90.0\ \text{mC}$

Exercise:

Problem:

In open heart surgery, a much smaller amount of energy will defibrillate the heart. (a) What voltage is applied to the $8.00\ \mu\text{F}$ capacitor of a heart defibrillator that stores 40.0 J of energy? (b) Find the amount of stored charge.

Solution:

(a) $3.16\ \text{kV}$

(b) $25.3\ \text{mC}$

Exercise:

Problem:

A $165\ \mu\text{F}$ capacitor is used in conjunction with a motor. How much energy is stored in it when 119 V is applied?

Exercise:

Problem:

Suppose you have a 9.00 V battery, a $2.00\ \mu\text{F}$ capacitor, and a $7.40\ \mu\text{F}$ capacitor. (a) Find the charge and energy stored if the capacitors are connected to the battery in series. (b) Do the same for a parallel connection.

Solution:

(a) $1.42 \times 10^{-5}\ \text{C}$, $6.38 \times 10^{-5}\ \text{J}$

(b) $8.46 \times 10^{-5}\ \text{C}$, $3.81 \times 10^{-4}\ \text{J}$

Exercise:

Problem:

A nervous physicist worries that the two metal shelves of his wood frame bookcase might obtain a high voltage if charged by static electricity, perhaps produced by friction. (a) What is the capacitance of the empty shelves if they have area $1.00 \times 10^2\ \text{m}^2$ and are 0.200 m apart? (b) What is the voltage between them if opposite charges of magnitude 2.00 nC are placed on them? (c) To show that this voltage poses a small hazard, calculate the energy stored.

Solution:

(a) $4.43 \times 10^{-12}\ \text{F}$

(b) $452\ \text{V}$

(c) $4.52 \times 10^{-7}\ \text{J}$

Exercise:

Problem:

Show that for a given dielectric material the maximum energy a parallel plate capacitor can store is directly proportional to the volume of dielectric (Volume $= A \cdot d$). Note that the applied voltage is limited by the dielectric strength.

Exercise:

Problem: Construct Your Own Problem

Consider a heart defibrillator similar to that discussed in [link]. Construct a problem in which you examine the charge stored in the capacitor of a defibrillator as a function of stored energy. Among the things to be considered are the applied voltage and whether it should vary with energy to be delivered, the range of energies involved, and the capacitance of the defibrillator. You may also wish to consider the much smaller energy needed for defibrillation during open-heart surgery as a variation on this problem.

Exercise:

Problem: Unreasonable Results

(a) On a particular day, it takes 9.60×10^3 J of electric energy to start a truck's engine. Calculate the capacitance of a capacitor that could store that amount of energy at 12.0 V. (b) What is unreasonable about this result? (c) Which assumptions are responsible?

Solution:

(a) 133 F

(b) Such a capacitor would be too large to carry with a truck. The size of the capacitor would be enormous.

(c) It is unreasonable to assume that a capacitor can store the amount of energy needed.

Glossary

defibrillator
: a machine used to provide an electrical shock to a heart attack victim's heart in order to restore the heart's normal rhythmic pattern

Introduction to Circuits and DC Instruments
class="introduction"

Electric circuits in a computer allow large amounts of data to be quickly and accurately analyzed.. (credit: Airman 1st Class Mike Meares, United States Air Force)



Electric circuits are commonplace. Some are simple, such as those in flashlights. Others, such as those used in supercomputers, are extremely complex.

This collection of modules takes the topic of electric circuits a step beyond simple circuits. When the circuit is purely resistive, everything in this module applies to both DC and AC. Matters become more complex when capacitance is involved. We do consider what happens when capacitors are connected to DC voltage sources, but the interaction of capacitors and other nonresistive devices with AC is left for a later chapter. Finally, a number of important DC instruments, such as meters that measure voltage and current, are covered in this chapter.

DC Voltmeters and Ammeters

- Explain why a voltmeter must be connected in parallel with the circuit.
- Draw a diagram showing an ammeter correctly connected in a circuit.
- Describe how a galvanometer can be used as either a voltmeter or an ammeter.
- Find the resistance that must be placed in series with a galvanometer to allow it to be used as a voltmeter with a given reading.
- Explain why measuring the voltage or current in a circuit can never be exact.

Voltmeters measure voltage, whereas **ammeters** measure current. Some of the meters in automobile dashboards, digital cameras, cell phones, and tuner-amplifiers are voltmeters or ammeters. (See [link].) The internal construction of the simplest of these meters and how they are connected to the system they monitor give further insight into applications of series and parallel connections.

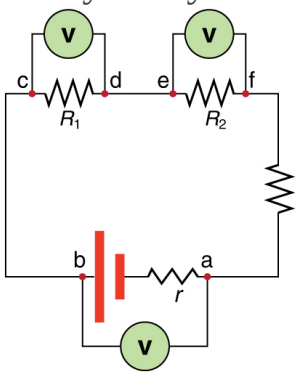


The fuel and temperature gauges (far right and far left, respectively) in this 1996 Volkswagen are voltmeters that register the voltage output of “sender” units, which are hopefully proportional to the amount of gasoline in the tank and the engine

temperature. (credit:
Christian Giersing)

Voltmeters are connected in parallel with whatever device's voltage is to be measured. A parallel connection is used because objects in parallel experience the same potential difference. (See [link], where the voltmeter is represented by the symbol V.)

Ammeters are connected in series with whatever device's current is to be measured. A series connection is used because objects in series have the same current passing through them. (See [link], where the ammeter is represented by the symbol A.)



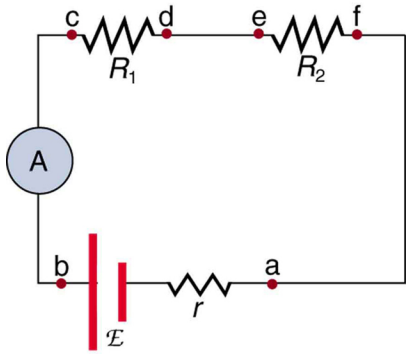
(a)



(b)

(a) To measure potential differences in this series circuit, the voltmeter (V) is placed in parallel with the

voltage source or either of the resistors. Note that terminal voltage is measured between points a and b. It is not possible to connect the voltmeter directly across the emf without including its internal resistance, r. (b) A digital voltmeter in use. (credit: Messtechniker, Wikimedia Commons)



An ammeter (A) is placed in series to measure current. All of the current in this circuit flows through the meter. The ammeter would have the same reading if located between points d and e or between points f and a as it does in the position shown. (Note that the script

capital E stands for emf, and r stands for the internal resistance of the source of potential difference.)

Analog Meters: Galvanometers

Analog meters have a needle that swivels to point at numbers on a scale, as opposed to **digital meters**, which have numerical readouts similar to a hand-held calculator. The heart of most analog meters is a device called a **galvanometer**, denoted by G. Current flow through a galvanometer, I_{G}, produces a proportional needle deflection. (This deflection is due to the force of a magnetic field upon a current-carrying wire.)

The two crucial characteristics of a given galvanometer are its resistance and current sensitivity. **Current sensitivity** is the current that gives a **full-scale deflection** of the galvanometer's needle, the maximum current that the instrument can measure. For example, a galvanometer with a current sensitivity of $50\ \mu\mathrm{A}$ has a maximum deflection of its needle when $50\ \mu\mathrm{A}$ flows through it, reads half-scale when $25\ \mu\mathrm{A}$ flows through it, and so on.

If such a galvanometer has a $25\text{-}\Omega$ resistance, then a voltage of only $V = \mathrm{IR} = (50\ \mu\mathrm{A})(25\ \Omega) = 1.25\ \mathrm{mV}$ produces a full-scale reading. By connecting resistors to this galvanometer in different ways, you can use it as either a voltmeter or ammeter that can measure a broad range of voltages or currents.

Galvanometer as Voltmeter

[link] shows how a galvanometer can be used as a voltmeter by connecting it in series with a large resistance, R. The value of the resistance R is

determined by the maximum voltage to be measured. Suppose you want 10 V to produce a full-scale deflection of a voltmeter containing a 25-Ω galvanometer with a 50-μA sensitivity. Then 10 V applied to the meter must produce a current of 50 μA. The total resistance must be
Equation:

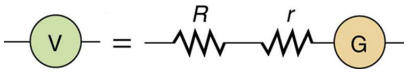
$$R_{\text{tot}} = R + r = \frac{V}{I} = \frac{10\text{ V}}{50\text{ μA}} = 200\text{ k }\Omega, \text{ or}$$

Equation:

$$R = R_{\text{tot}} - r = 200\text{ k}\Omega - 25\text{ }\Omega \approx 200\text{ k}\Omega.$$

(R is so large that the galvanometer resistance, r, is nearly negligible.) Note that 5 V applied to this voltmeter produces a half-scale deflection by producing a 25-μA current through the meter, and so the voltmeter's reading is proportional to voltage as desired.

This voltmeter would not be useful for voltages less than about half a volt, because the meter deflection would be small and difficult to read accurately. For other voltage ranges, other resistances are placed in series with the galvanometer. Many meters have a choice of scales. That choice involves switching an appropriate resistance into series with the galvanometer.



A large resistance R placed in series with a galvanometer G produces a voltmeter, the full-scale deflection of which depends on the choice of R.

The larger the voltage to be measured, the larger R must be. (Note that r represents the internal resistance of the galvanometer.)

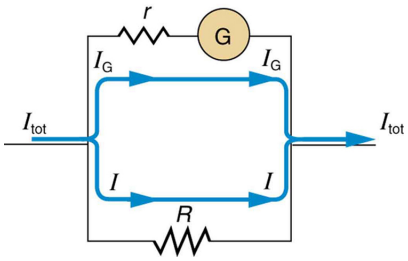
Galvanometer as Ammeter

The same galvanometer can also be made into an ammeter by placing it in parallel with a small resistance R, often called the **shunt resistance**, as shown in [link]. Since the shunt resistance is small, most of the current passes through it, allowing an ammeter to measure currents much greater than those producing a full-scale deflection of the galvanometer.

Suppose, for example, an ammeter is needed that gives a full-scale deflection for 1.0 A, and contains the same $25\text{-}\Omega$ galvanometer with its $50\text{-}\mu\text{A}$ sensitivity. Since R and r are in parallel, the voltage across them is the same.

These IR drops are $\text{IR} = I_{\text{G}}r$ so that $\text{IR} = \frac{I_{\text{G}}}{I} = \frac{R}{r}$. Solving for R, and noting that I_{G} is $50\ \mu\text{A}$ and I is 0.999950 A, we have
Equation:

$$R = r\frac{I_{\text{G}}}{I} = (25\,\Omega)\frac{50\ \mu\text{A}}{0.999950\ \text{A}} = 1.25\times10^{-3}\,\Omega.$$



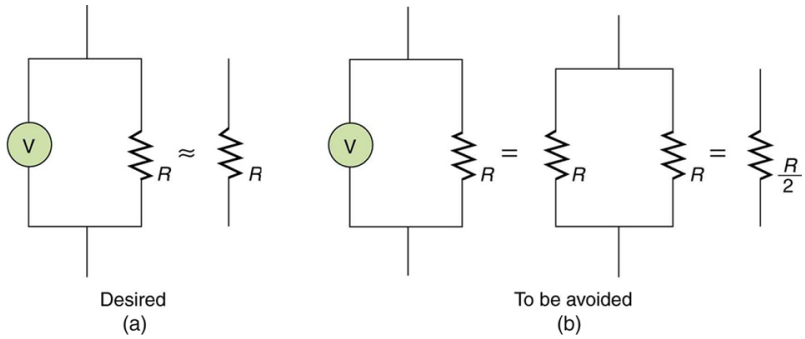
A small shunt resistance R placed in parallel with a galvanometer G produces an ammeter, the full-scale deflection of which depends on the choice of R. The larger the current to be measured, the smaller R must be. Most of the current (I) flowing through the meter is shunted through R to protect the galvanometer. (Note that r represents the internal resistance of the galvanometer.) Ammeters may also have multiple scales for greater flexibility in application. The various scales are

achieved by switching various shunt resistances in parallel with the galvanometer—the greater the maximum current to be measured, the smaller the shunt resistance must be.

Taking Measurements Alters the Circuit

When you use a voltmeter or ammeter, you are connecting another resistor to an existing circuit and, thus, altering the circuit. Ideally, voltmeters and ammeters do not appreciably affect the circuit, but it is instructive to examine the circumstances under which they do or do not interfere.

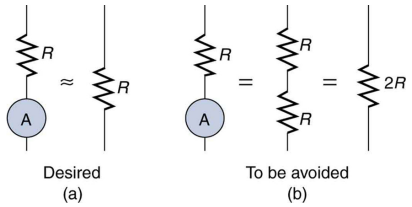
First, consider the voltmeter, which is always placed in parallel with the device being measured. Very little current flows through the voltmeter if its resistance is a few orders of magnitude greater than the device, and so the circuit is not appreciably affected. (See [link](a).) (A large resistance in parallel with a small one has a combined resistance essentially equal to the small one.) If, however, the voltmeter's resistance is comparable to that of the device being measured, then the two in parallel have a smaller resistance, appreciably affecting the circuit. (See [link](b).) The voltage across the device is not the same as when the voltmeter is out of the circuit.



(a) A voltmeter having a resistance much larger than the device ($R_{\text{Voltmeter}} >> R$) with which it is in parallel produces a parallel resistance essentially the same as the device and does not appreciably affect the circuit being measured. (b) Here the voltmeter has the same resistance as the device ($R_{\text{Voltmeter}} \cong R$), so that the parallel resistance is half of what it is when the voltmeter is not connected. This is an example of a significant alteration of the circuit and is to be avoided.

An ammeter is placed in series in the branch of the circuit being measured, so that its resistance adds to that branch. Normally, the ammeter's resistance is very small compared with the resistances of the devices in the circuit, and so the extra resistance is negligible. (See [link](a).) However, if very small load resistances are involved, or if the ammeter is not as low in resistance as it should be, then the total series resistance is significantly greater, and the current in the branch being measured is reduced. (See [link](b).)

A practical problem can occur if the ammeter is connected incorrectly. If it was put in parallel with the resistor to measure the current in it, you could possibly damage the meter; the low resistance of the ammeter would allow most of the current in the circuit to go through the galvanometer, and this current would be larger since the effective resistance is smaller.



(a) An ammeter normally has such a small resistance that the total series resistance in the branch being measured is not appreciably increased. The circuit is essentially unaltered compared with when the ammeter is absent. (b) Here the ammeter’s resistance is the same as that of the branch, so that the total resistance is doubled and the current is half what it is without the ammeter. This significant alteration of the circuit is to be avoided.

One solution to the problem of voltmeters and ammeters interfering with the circuits being measured is to use galvanometers with greater sensitivity.

This allows construction of voltmeters with greater resistance and ammeters with smaller resistance than when less sensitive galvanometers are used.

There are practical limits to galvanometer sensitivity, but it is possible to get analog meters that make measurements accurate to a few percent. Note that the inaccuracy comes from altering the circuit, not from a fault in the meter.

Note:
Connections: Limits to Knowledge
Making a measurement alters the system being measured in a manner that produces uncertainty in the measurement. For macroscopic systems, such as the circuits discussed in this module, the alteration can usually be made negligibly small, but it cannot be eliminated entirely. For submicroscopic systems, such as atoms, nuclei, and smaller particles, measurement alters the system in a manner that cannot be made arbitrarily small. This actually limits knowledge of the system—even limiting what nature can know about itself. We shall see profound implications of this when the Heisenberg uncertainty principle is discussed in the modules on quantum mechanics.
There is another measurement technique based on drawing no current at all and, hence, not altering the circuit at all. These are called null measurements and are the topic of Null Measurements. Digital meters that employ solid-state electronics and null measurements can attain accuracies of one part in 10^6.

Exercise:
Check Your Understanding

Problem:

Digital meters are able to detect smaller currents than analog meters employing galvanometers. How does this explain their ability to measure voltage and current more accurately than analog meters?

Solution:

Since digital meters require less current than analog meters, they alter the circuit less than analog meters. Their resistance as a voltmeter can be far greater than an analog meter, and their resistance as an ammeter can be far less than an analog meter. Consult [link] and [link] and their discussion in the text.

Note:
PhET Explorations: Circuit Construction Kit (DC Only), Virtual Lab
Stimulate a neuron and monitor what happens. Pause, rewind, and move forward in time in order to observe the ions as they move across the neuron membrane.

Circuit Construction Kit (DC Only), Virtual Lab

Section Summary

- Voltmeters measure voltage, and ammeters measure current.
- A voltmeter is placed in parallel with the voltage source to receive full voltage and must have a large resistance to limit its effect on the circuit.
- An ammeter is placed in series to get the full current flowing through a branch and must have a small resistance to limit its effect on the circuit.

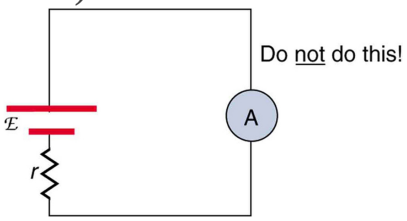
- Both can be based on the combination of a resistor and a galvanometer, a device that gives an analog reading of current.
- Standard voltmeters and ammeters alter the circuit being measured and are thus limited in accuracy.

Conceptual Questions

Exercise:

Problem:

Why should you not connect an ammeter directly across a voltage source as shown in [link]? (Note that script E in the figure stands for emf.)



Exercise:

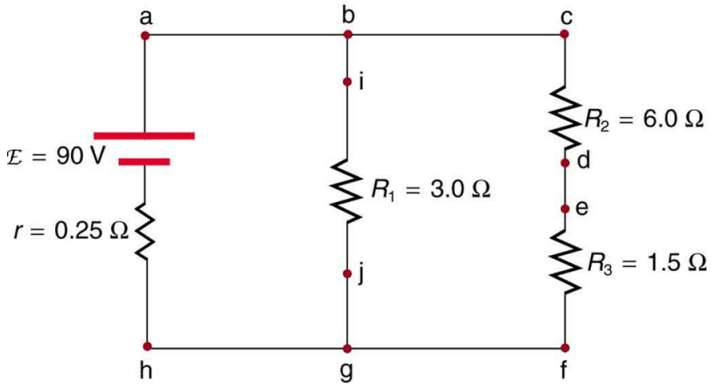
Problem:

Suppose you are using a multimeter (one designed to measure a range of voltages, currents, and resistances) to measure current in a circuit and you inadvertently leave it in a voltmeter mode. What effect will the meter have on the circuit? What would happen if you were measuring voltage but accidentally put the meter in the ammeter mode?

Exercise:

Problem:

Specify the points to which you could connect a voltmeter to measure the following potential differences in [link]: (a) the potential difference of the voltage source; (b) the potential difference across R_1; (c) across R_2; (d) across R_3; (e) across R_2 and R_3. Note that there may be more than one answer to each part.



Exercise:

Problem:

To measure currents in [link], you would replace a wire between two points with an ammeter. Specify the points between which you would place an ammeter to measure the following: (a) the total current; (b) the current flowing through R_1; (c) through R_2; (d) through R_3. Note that there may be more than one answer to each part.

Problem Exercises

Exercise:

Problem:

What is the sensitivity of the galvanometer (that is, what current gives a full-scale deflection) inside a voltmeter that has a 1.00-MΩ resistance on its 30.0-V scale?

Solution:

30 μA

Exercise:

Problem:

What is the sensitivity of the galvanometer (that is, what current gives a full-scale deflection) inside a voltmeter that has a $25.0\text{-k}\Omega$ resistance on its 100-V scale?

Exercise:

Problem:

Find the resistance that must be placed in series with a $25.0\text{-}\Omega$ galvanometer having a $50.0-\mu\text{A}$ sensitivity (the same as the one discussed in the text) to allow it to be used as a voltmeter with a 0.100-V full-scale reading.

Solution:

$1.98\ \text{k}\Omega$

Exercise:

Problem:

Find the resistance that must be placed in series with a $25.0\text{-}\Omega$ galvanometer having a $50.0\text{-}\mu\text{A}$ sensitivity (the same as the one discussed in the text) to allow it to be used as a voltmeter with a 3000-V full-scale reading. Include a circuit diagram with your solution.

Exercise:

Problem:

Find the resistance that must be placed in parallel with a $25.0\text{-}\Omega$ galvanometer having a $50.0\text{-}\mu\text{A}$ sensitivity (the same as the one discussed in the text) to allow it to be used as an ammeter with a 10.0-A full-scale reading. Include a circuit diagram with your solution.

Solution:
Equation:

$$1.25\times10^{-4}\ \Omega$$

Exercise:

Problem:

Find the resistance that must be placed in parallel with a $25.0\text{-}\Omega$ galvanometer having a $50.0\text{-}\mu\text{A}$ sensitivity (the same as the one discussed in the text) to allow it to be used as an ammeter with a 300-mA full-scale reading.

Exercise:

Problem:

Find the resistance that must be placed in series with a $10.0\text{-}\Omega$ galvanometer having a $100\text{-}\mu\text{A}$ sensitivity to allow it to be used as a voltmeter with: (a) a 300-V full-scale reading, and (b) a 0.300-V full-scale reading.

Solution:

(a) $3.00\ \text{M}\Omega$

(b) $2.99\ \text{k}\Omega$

Exercise:

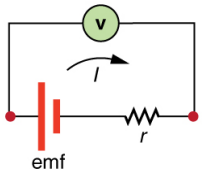
Problem:

Find the resistance that must be placed in parallel with a $10.0\text{-}\Omega$ galvanometer having a $100\text{-}\mu\text{A}$ sensitivity to allow it to be used as an ammeter with: (a) a 20.0-A full-scale reading, and (b) a 100-mA full-scale reading.

Exercise:

Problem:

Suppose you measure the terminal voltage of a 1.585-V alkaline cell having an internal resistance of $0.100\,\Omega$ by placing a $1.00\text{-k}\Omega$ voltmeter across its terminals. (See [link].) (a) What current flows? (b) Find the terminal voltage. (c) To see how close the measured terminal voltage is to the emf, calculate their ratio.



Solution:

(a) 1.58 mA

(b) 1.5848 V (need four digits to see the difference)

(c) 0.99990 (need five digits to see the difference from unity)

Exercise:

Problem:

Suppose you measure the terminal voltage of a 3.200-V lithium cell having an internal resistance of $5.00\,\Omega$ by placing a $1.00\text{-k}\Omega$ voltmeter across its terminals. (a) What current flows? (b) Find the terminal voltage. (c) To see how close the measured terminal voltage is to the emf, calculate their ratio.

Exercise:

Problem:

A certain ammeter has a resistance of $5.00\times10^{-5}\ \Omega$ on its 3.00-A scale and contains a $10.0\text{-}\Omega$ galvanometer. What is the sensitivity of the galvanometer?

Solution:

$15.0\ \mu\text{A}$

Exercise:

Problem:

A $1.00\text{-M}\Omega$ voltmeter is placed in parallel with a $75.0\text{-k}\Omega$ resistor in a circuit. (a) Draw a circuit diagram of the connection. (b) What is the resistance of the combination? (c) If the voltage across the combination is kept the same as it was across the $75.0\text{-k}\Omega$ resistor alone, what is the percent increase in current? (d) If the current through the combination is kept the same as it was through the $75.0\text{-k}\Omega$ resistor alone, what is the percentage decrease in voltage? (e) Are the changes found in parts (c) and (d) significant? Discuss.

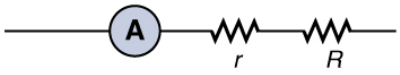
Exercise:

Problem:

A $0.0200\text{-}\Omega$ ammeter is placed in series with a $10.00\text{-}\Omega$ resistor in a circuit. (a) Draw a circuit diagram of the connection. (b) Calculate the resistance of the combination. (c) If the voltage is kept the same across the combination as it was through the $10.00\text{-}\Omega$ resistor alone, what is the percent decrease in current? (d) If the current is kept the same through the combination as it was through the $10.00\text{-}\Omega$ resistor alone, what is the percent increase in voltage? (e) Are the changes found in parts (c) and (d) significant? Discuss.

Solution:

(a)



(b) $10.02\ \Omega$

(c) 0.9980, or a 2.0×10^{-1} percent decrease

(d) 1.002, or a 2.0×10^{-1} percent increase

(e) Not significant.

Exercise:

Problem: Unreasonable Results

Suppose you have a $40.0\text{-}\Omega$ galvanometer with a $25.0\text{-}\mu\text{A}$ sensitivity. (a) What resistance would you put in series with it to allow it to be used as a voltmeter that has a full-scale deflection for 0.500 mV? (b) What is unreasonable about this result? (c) Which assumptions are responsible?

Exercise:

Problem: Unreasonable Results

(a) What resistance would you put in parallel with a $40.0\text{-}\Omega$ galvanometer having a $25.0\text{-}\mu\text{A}$ sensitivity to allow it to be used as an ammeter that has a full-scale deflection for $10.0\text{-}\mu\text{A}$? (b) What is unreasonable about this result? (c) Which assumptions are responsible?

Solution:

(a) $-66.7\ \Omega$

(b) You can't have negative resistance.

(c) It is unreasonable that I_{G} is greater than I_{tot} (see [link]). You cannot achieve a full-scale deflection using a current less than the sensitivity of the galvanometer.

Glossary

voltmeter
: an instrument that measures voltage

ammeter

an instrument that measures current

analog meter
: a measuring instrument that gives a readout in the form of a needle movement over a marked gauge

digital meter
: a measuring instrument that gives a readout in a digital form

galvanometer
: an analog measuring device, denoted by G, that measures current flow using a needle deflection caused by a magnetic field force acting upon a current-carrying wire

current sensitivity
: the maximum current that a galvanometer can read

full-scale deflection
: the maximum deflection of a galvanometer needle, also known as current sensitivity; a galvanometer with a full-scale deflection of $50\ \mu\text{A}$ has a maximum deflection of its needle when $50\ \mu\text{A}$ flows through it

shunt resistance
: a small resistance R placed in parallel with a galvanometer G to produce an ammeter; the larger the current to be measured, the smaller R must be; most of the current flowing through the meter is shunted through R to protect the galvanometer

Electromotive Force: Terminal Voltage

- Compare and contrast the voltage and the electromagnetic force of an electric power source.
- Describe what happens to the terminal voltage, current, and power delivered to a load as internal resistance of the voltage source increases (due to aging of batteries, for example).
- Explain why it is beneficial to use more than one voltage source connected in parallel.

When you forget to turn off your car lights, they slowly dim as the battery runs down. Why don't they simply blink off when the battery's energy is gone? Their gradual dimming implies that battery output voltage decreases as the battery is depleted.

Furthermore, if you connect an excessive number of 12-V lights in parallel to a car battery, they will be dim even when the battery is fresh and even if the wires to the lights have very low resistance. This implies that the battery's output voltage is reduced by the overload.

The reason for the decrease in output voltage for depleted or overloaded batteries is that all voltage sources have two fundamental parts—a source of electrical energy and an **internal resistance**. Let us examine both.

Electromotive Force

You can think of many different types of voltage sources. Batteries themselves come in many varieties. There are many types of mechanical/electrical generators, driven by many different energy sources, ranging from nuclear to wind. Solar cells create voltages directly from light, while thermoelectric devices create voltage from temperature differences.

A few voltage sources are shown in [link]. All such devices create a **potential difference** and can supply current if connected to a resistance. On the small scale, the potential difference creates an electric field that exerts force on charges, causing current. We thus use the name **electromotive force**, abbreviated emf.

Emf is not a force at all; it is a special type of potential difference. To be precise, the electromotive force (emf) is the potential difference of a source when no current is flowing. Units of emf are volts.



A variety of voltage sources (clockwise from top left): the Brazos Wind Farm in Fluvanna, Texas (credit: Leaflet, Wikimedia Commons); the Krasnoyarsk Dam in Russia (credit: Alex Polezhaev); a solar farm (credit: U.S. Department of Energy); and a group of nickel metal hydride batteries (credit: Tiaa Monto). The voltage output of each depends on its construction and load, and equals emf only if there is no load.

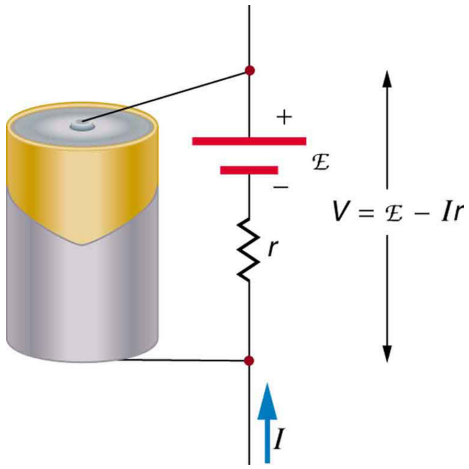
Electromotive force is directly related to the source of potential difference, such as the particular combination of chemicals in a battery. However, emf

differs from the voltage output of the device when current flows. The voltage across the terminals of a battery, for example, is less than the emf when the battery supplies current, and it declines further as the battery is depleted or loaded down. However, if the device's output voltage can be measured without drawing current, then output voltage will equal emf (even for a very depleted battery).

Internal Resistance

As noted before, a 12-V truck battery is physically larger, contains more charge and energy, and can deliver a larger current than a 12-V motorcycle battery. Both are lead-acid batteries with identical emf, but, because of its size, the truck battery has a smaller internal resistance r. Internal resistance is the inherent resistance to the flow of current within the source itself.

[link] is a schematic representation of the two fundamental parts of any voltage source. The emf (represented by a script E in the figure) and internal resistance r are in series. The smaller the internal resistance for a given emf, the more current and the more power the source can supply.

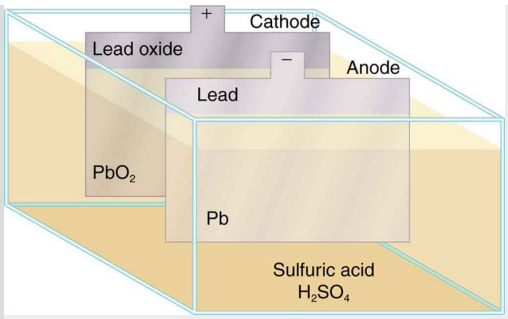


Any voltage source (in this case, a carbon-zinc dry cell) has an emf related to its source of

potential difference, and an internal resistance r related to its construction. (Note that the script E stands for emf.). Also shown are the output terminals across which the terminal voltage V is measured. Since $V = \text{emf} - \text{Ir}$, terminal voltage equals emf only if there is no current flowing.

The internal resistance r can behave in complex ways. As noted, r increases as a battery is depleted. But internal resistance may also depend on the magnitude and direction of the current through a voltage source, its temperature, and even its history. The internal resistance of rechargeable nickel-cadmium cells, for example, depends on how many times and how deeply they have been depleted.

Note:
Things Great and Small: The Submicroscopic Origin of Battery Potential
Various types of batteries are available, with emfs determined by the combination of chemicals involved. We can view this as a molecular reaction (what much of chemistry is about) that separates charge.
The lead-acid battery used in cars and other vehicles is one of the most common types. A single cell (one of six) of this battery is seen in [link].
The cathode (positive) terminal of the cell is connected to a lead oxide plate, while the anode (negative) terminal is connected to a lead plate. Both plates are immersed in sulfuric acid, the electrolyte for the system.

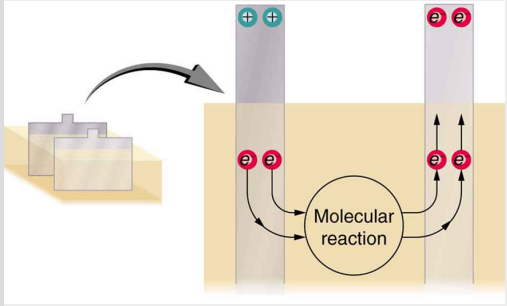


Artist’s conception of a lead-acid cell. Chemical reactions in a lead-acid cell separate charge, sending negative charge to the anode, which is connected to the lead plates. The lead oxide plates are connected to the positive or cathode terminal of the cell. Sulfuric acid conducts the charge as well as participating in the chemical reaction.

The details of the chemical reaction are left to the reader to pursue in a chemistry text, but their results at the molecular level help explain the potential created by the battery. [link] shows the result of a single chemical reaction. Two electrons are placed on the anode, making it negative, provided that the cathode supplied two electrons. This leaves the cathode positively charged, because it has lost two electrons. In short, a separation of charge has been driven by a chemical reaction.

Note that the reaction will not take place unless there is a complete circuit to allow two electrons to be supplied to the cathode. Under many circumstances, these electrons come from the anode, flow through a resistance, and return to the cathode. Note also that since the chemical

reactions involve substances with resistance, it is not possible to create the emf without an internal resistance.



Artist’s conception of two electrons being forced onto the anode of a cell and two electrons being removed from the cathode of the cell. The chemical reaction in a lead-acid battery places two electrons on the anode and removes two from the cathode. It requires a closed circuit to proceed, since the two electrons must be supplied to the cathode.

Why are the chemicals able to produce a unique potential difference? Quantum mechanical descriptions of molecules, which take into account the types of atoms and numbers of electrons in them, are able to predict the energy states they can have and the energies of reactions between them.

In the case of a lead-acid battery, an energy of 2 eV is given to each electron sent to the anode. Voltage is defined as the electrical potential

energy divided by charge: $V = \frac{P_{\text{E}}}{q}$. An electron volt is the energy given to a single electron by a voltage of 1 V. So the voltage here is 2 V, since 2 eV is given to each electron. It is the energy produced in each molecular reaction that produces the voltage. A different reaction produces a different energy and, hence, a different voltage.

Terminal Voltage

The voltage output of a device is measured across its terminals and, thus, is called its **terminal voltage** V. Terminal voltage is given by
Equation:

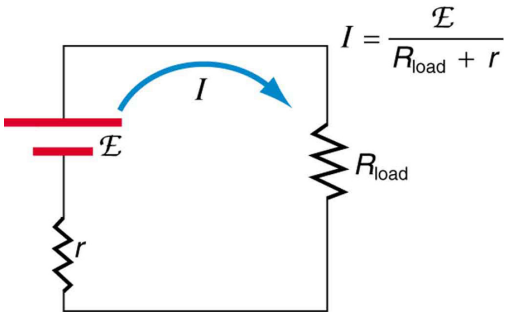
$$V = \text{emf} - \text{Ir},$$

where r is the internal resistance and I is the current flowing at the time of the measurement.

I is positive if current flows away from the positive terminal, as shown in [link]. You can see that the larger the current, the smaller the terminal voltage. And it is likewise true that the larger the internal resistance, the smaller the terminal voltage.

Suppose a load resistance R_{load} is connected to a voltage source, as in [link]. Since the resistances are in series, the total resistance in the circuit is $R_{\text{load}} + r$. Thus the current is given by Ohm's law to be
Equation:

$$I = \frac{\text{emf}}{R_{\text{load}} + r}.$$



Schematic of a voltage source and its load R_{load}. Since the internal resistance r is in series with the load, it can significantly affect the terminal voltage and current delivered to the load. (Note that the script E stands for emf.)

We see from this expression that the smaller the internal resistance r, the greater the current the voltage source supplies to its load R_{load}. As batteries are depleted, r increases. If r becomes a significant fraction of the load resistance, then the current is significantly reduced, as the following example illustrates.

Example:
Calculating Terminal Voltage, Power Dissipation, Current, and Resistance: Terminal Voltage and Load
A certain battery has a 12.0-V emf and an internal resistance of 0.100 Ω. (a) Calculate its terminal voltage when connected to a 10.0-Ω load. (b) What is the terminal voltage when connected to a 0.500-Ω load? (c) What power does the 0.500-Ω load dissipate? (d) If the internal resistance grows

to $0.500\ \Omega$, find the current, terminal voltage, and power dissipated by a 0.500-Ω load.

Strategy

The analysis above gave an expression for current when internal resistance is taken into account. Once the current is found, the terminal voltage can be calculated using the equation $V = \text{emf} - \text{Ir}$. Once current is found, the power dissipated by a resistor can also be found.

Solution for (a)

Entering the given values for the emf, load resistance, and internal resistance into the expression above yields

Equation:

$$I = \frac{\text{emf}}{R_{\text{load}} + r} = \frac{12.0\ \text{V}}{10.1\,\Omega} = 1.188\ \text{A}.$$

Enter the known values into the equation $V = \text{emf} - \text{Ir}$ to get the terminal voltage:

Equation:

$$\begin{aligned} V &= \text{emf} - Ir = 12.0\ \text{V} - (1.188\ \text{A})(0.100\ \Omega) \\ &= 11.9\ \text{V}. \end{aligned}$$

Discussion for (a)

The terminal voltage here is only slightly lower than the emf, implying that $10.0\ \Omega$ is a light load for this particular battery.

Solution for (b)

Similarly, with $R_{\text{load}} = 0.500\,\Omega$, the current is

Equation:

$$I = \frac{\text{emf}}{R_{\text{load}} + r} = \frac{12.0\ \text{V}}{0.600\,\Omega} = 20.0\ \text{A}.$$

The terminal voltage is now

Equation:

$$\begin{aligned} V &= \text{emf} - Ir = 12.0\ \text{V} - (20.0\ \text{A})(0.100\ \Omega) \\ &= 10.0\ \text{V}. \end{aligned}$$

Discussion for (b)
This terminal voltage exhibits a more significant reduction compared with emf, implying $0.500\,\Omega$ is a heavy load for this battery.
Solution for (c)
The power dissipated by the 0.500 - Ω load can be found using the formula $P = I^2R$. Entering the known values gives
Equation:

$$P_{\text{load}} = I^2R_{\text{load}} = (20.0\text{ A})^2(0.500\,\Omega) = 2.00 \times 10^2\text{ W}.$$

Discussion for (c)
Note that this power can also be obtained using the expressions $\frac{V^2}{R}$ or IV, where V is the terminal voltage (10.0 V in this case).
Solution for (d)
Here the internal resistance has increased, perhaps due to the depletion of the battery, to the point where it is as great as the load resistance. As before, we first find the current by entering the known values into the expression, yielding
Equation:

$$I = \frac{\text{emf}}{R_{\text{load}} + r} = \frac{12.0\text{ V}}{1.00\,\Omega} = 12.0\text{ A}.$$

Now the terminal voltage is
Equation:

$$\begin{aligned} V &= \text{emf} - Ir = 12.0\text{ V} - (12.0\text{ A})(0.500\,\Omega) \\ &= 6.00\text{ V}, \end{aligned}$$

and the power dissipated by the load is
Equation:

$$P_{\text{load}} = I^2R_{\text{load}} = (12.0\text{ A})^2(0.500\,\Omega) = 72.0\text{ W}.$$

Discussion for (d)
We see that the increased internal resistance has significantly decreased terminal voltage, current, and power delivered to a load.

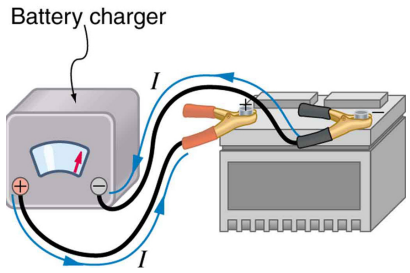
Battery testers, such as those in [link], use small load resistors to intentionally draw current to determine whether the terminal voltage drops below an acceptable level. They really test the internal resistance of the battery. If internal resistance is high, the battery is weak, as evidenced by its low terminal voltage.



These two battery testers measure terminal voltage under a load to determine the condition of a battery. The large device is being used by a U.S. Navy electronics technician to test large batteries aboard the aircraft carrier USS *Nimitz* and has a small resistance that can dissipate large amounts of power. (credit: U.S. Navy photo by Photographer's Mate Airman Jason A. Johnston) The small device is used on small batteries and has a digital display to indicate the acceptability of their terminal voltage. (credit: Keith Williamson)

Some batteries can be recharged by passing a current through them in the direction opposite to the current they supply to a resistance. This is done routinely in cars and batteries for small electrical appliances and electronic devices, and is represented pictorially in [link]. The voltage output of the battery charger must be greater than the emf of the battery to reverse current

through it. This will cause the terminal voltage of the battery to be greater than the emf, since $V = \text{emf} - \text{Ir}$, and I is now negative.



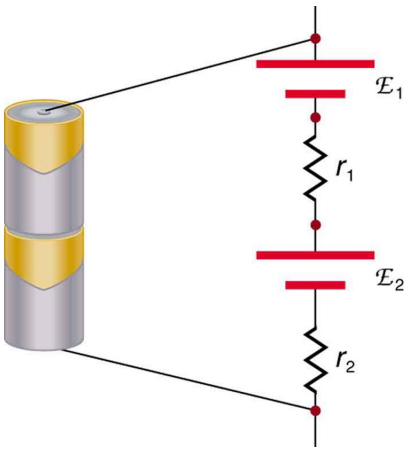
A car battery charger reverses the normal direction of current through a battery, reversing its chemical reaction and replenishing its chemical potential.

Multiple Voltage Sources

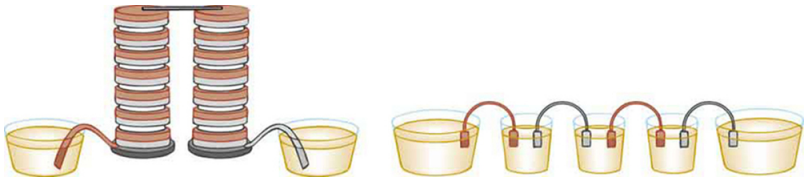
There are two voltage sources when a battery charger is used. Voltage sources connected in series are relatively simple. When voltage sources are in series, their internal resistances add and their emfs add algebraically. (See [link].) Series connections of voltage sources are common—for example, in flashlights, toys, and other appliances. Usually, the cells are in series in order to produce a larger total emf.

But if the cells oppose one another, such as when one is put into an appliance backward, the total emf is less, since it is the algebraic sum of the individual emfs.

A battery is a multiple connection of voltaic cells, as shown in [link]. The disadvantage of series connections of cells is that their internal resistances add. One of the authors once owned a 1957 MGA that had two 6-V batteries in series, rather than a single 12-V battery. This arrangement produced a large internal resistance that caused him many problems in starting the engine.

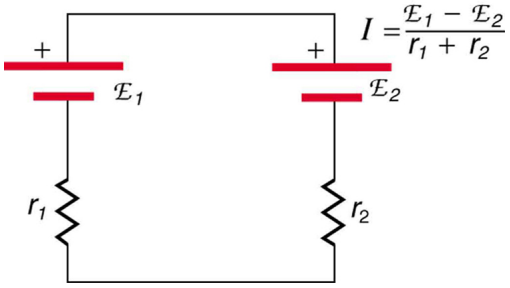


A series connection of two voltage sources. The emfs (each labeled with a script E) and internal resistances add, giving a total emf of $\text{emf}_1 + \text{emf}_2$ and a total internal resistance of $r_1 + r_2$.



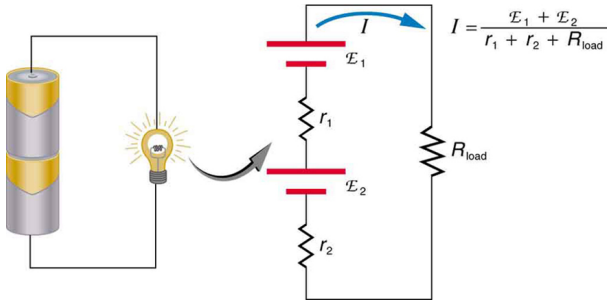
Batteries are multiple connections of individual cells, as shown in this modern rendition of an old print. Single cells, such as AA or C cells, are commonly called batteries, although this is technically incorrect.

If the *series* connection of two voltage sources is made into a complete circuit with the emfs in opposition, then a current of magnitude $I = \frac{(\text{emf}_1 - \text{emf}_2)}{r_1 + r_2}$ flows. See [link], for example, which shows a circuit exactly analogous to the battery charger discussed above. If two voltage sources in series with emfs in the same sense are connected to a load R_{load}, as in [link], then $I = \frac{(\text{emf}_1 + \text{emf}_2)}{r_1 + r_2 + R_{\text{load}}}$ flows.



These two voltage sources are connected in series with their emfs in opposition. Current flows in the direction of the greater emf and is limited

to $I = \frac{(\text{emf}_1 - \text{emf}_2)}{r_1 + r_2}$ by the sum of the internal resistances. (Note that each emf is represented by script E in the figure.) A battery charger connected to a battery is an example of such a connection. The charger must have a larger emf than the battery to reverse current through it.

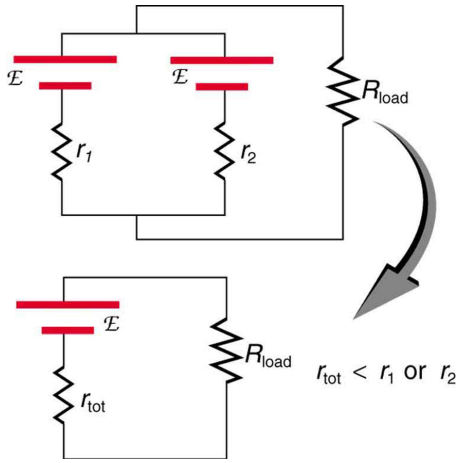


This schematic represents a flashlight with two cells (voltage sources) and a single bulb (load resistance) in series. The current that flows is $I = \frac{(\text{emf}_1 + \text{emf}_2)}{r_1 + r_2 + R_{\text{load}}}$. (Note that each emf is represented by script E in the figure.)

Note:
Take-Home Experiment: Flashlight Batteries
Find a flashlight that uses several batteries and find new and old batteries. Based on the discussions in this module, predict the brightness of the flashlight when different combinations of batteries are used. Do your predictions match what you observe? Now place new batteries in the flashlight and leave the flashlight switched on for several hours. Is the flashlight still quite bright? Do the same with the old batteries. Is the flashlight as bright when left on for the same length of time with old and new batteries? What does this say for the case when you are limited in the number of available new batteries?

[link] shows two voltage sources with identical emfs in parallel and connected to a load resistance. In this simple case, the total emf is the same as the individual emfs. But the total internal resistance is reduced, since the internal resistances are in parallel. The parallel connection thus can produce a larger current.

Here, $I = \frac{\text{emf}}{(r_{\text{tot}} + R_{\text{load}})}$ flows through the load, and r_{tot} is less than those of the individual batteries. For example, some diesel-powered cars use two 12-V batteries in parallel; they produce a total emf of 12 V but can deliver the larger current needed to start a diesel engine.



Two voltage sources with identical emfs (each labeled by script E) connected in parallel produce the same emf but have a smaller total internal resistance than the individual sources. Parallel combinations are often used to deliver more current. Here $I = \frac{\text{emf}}{(r_{\text{tot}} + R_{\text{load}})}$ flows through the load.

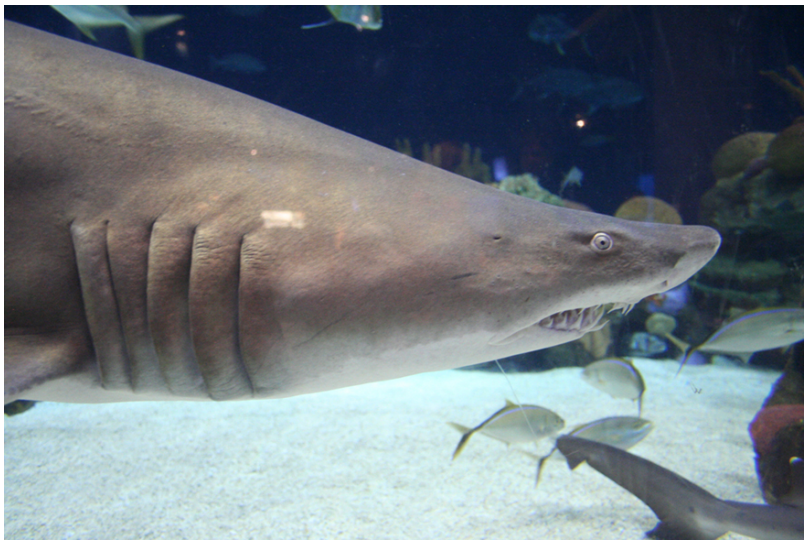
Animals as Electrical Detectors

A number of animals both produce and detect electrical signals. Fish, sharks, platypuses, and echidnas (spiny anteaters) all detect electric fields generated by nerve activity in prey. Electric eels produce their own emf through biological cells (electric organs) called electroplaques, which are arranged in both series and parallel as a set of batteries.

Electroplaques are flat, disk-like cells; those of the electric eel have a voltage of 0.15 V across each one. These cells are usually located toward the head or tail of the animal, although in the case of the electric eel, they are found along the entire body. The electroplaques in the South American eel are arranged in 140 rows, with each row stretching horizontally along the body and containing 5,000 electroplaques. This can yield an emf of approximately 600 V, and a current of 1 A—deadly.

The mechanism for detection of external electric fields is similar to that for producing nerve signals in the cell through depolarization and

repolarization—the movement of ions across the cell membrane. Within the fish, weak electric fields in the water produce a current in a gel-filled canal that runs from the skin to sensing cells, producing a nerve signal. The Australian platypus, one of the very few mammals that lay eggs, can detect fields of 30 $\frac{\text{mV}}{\text{m}}$, while sharks have been found to be able to sense a field in their snouts as small as 100 $\frac{\text{mV}}{\text{m}}$ ([link]). Electric eels use their own electric fields produced by the electroplaques to stun their prey or enemies.



Sand tiger sharks (*Carcharias taurus*), like this one at the Minnesota Zoo, use electroreceptors in their snouts to locate prey. (credit: Jim Winstead, Flickr)

Solar Cell Arrays

Another example dealing with multiple voltage sources is that of combinations of solar cells—wired in both series and parallel combinations to yield a desired voltage and current. Photovoltaic generation (PV), the conversion of sunlight directly into electricity, is based upon the

photoelectric effect, in which photons hitting the surface of a solar cell create an electric current in the cell.

Most solar cells are made from pure silicon—either as single-crystal silicon, or as a thin film of silicon deposited upon a glass or metal backing. Most single cells have a voltage output of about 0.5 V, while the current output is a function of the amount of sunlight upon the cell (the incident solar radiation—the insolation). Under bright noon sunlight, a current of about $100\ \mathrm{mA/cm^2}$ of cell surface area is produced by typical single-crystal cells.

Individual solar cells are connected electrically in modules to meet electrical-energy needs. They can be wired together in series or in parallel —connected like the batteries discussed earlier. A solar-cell array or module usually consists of between 36 and 72 cells, with a power output of 50 W to 140 W.

The output of the solar cells is direct current. For most uses in a home, AC is required, so a device called an inverter must be used to convert the DC to AC. Any extra output can then be passed on to the outside electrical grid for sale to the utility.

Note:
Take-Home Experiment: Virtual Solar Cells
One can assemble a “virtual” solar cell array by using playing cards, or business or index cards, to represent a solar cell. Combinations of these cards in series and/or parallel can model the required array output. Assume each card has an output of 0.5 V and a current (under bright light) of 2 A. Using your cards, how would you arrange them to produce an output of 6 A at 3 V (18 W)?
Suppose you were told that you needed only 18 W (but no required voltage). Would you need more cards to make this arrangement?

Section Summary

- All voltage sources have two fundamental parts—a source of electrical energy that has a characteristic electromotive force (emf), and an internal resistance r.
- The emf is the potential difference of a source when no current is flowing.
- The numerical value of the emf depends on the source of potential difference.
- The internal resistance r of a voltage source affects the output voltage when a current flows.
- The voltage output of a device is called its terminal voltage V and is given by $V = \text{emf} - \text{Ir}$, where I is the electric current and is positive when flowing away from the positive terminal of the voltage source.
- When multiple voltage sources are in series, their internal resistances add and their emfs add algebraically.
- Solar cells can be wired in series or parallel to provide increased voltage or current, respectively.

Conceptual Questions

Exercise:

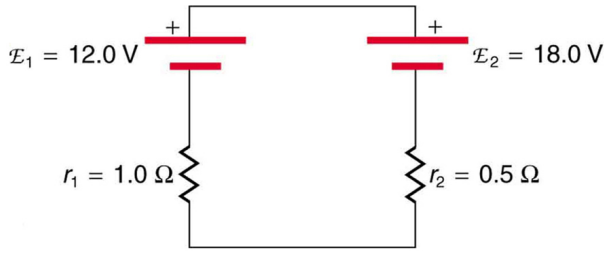
Problem:

Is every emf a potential difference? Is every potential difference an emf? Explain.

Exercise:

Problem:

Explain which battery is doing the charging and which is being charged in [link].



Exercise:

Problem:

Given a battery, an assortment of resistors, and a variety of voltage and current measuring devices, describe how you would determine the internal resistance of the battery.

Exercise:

Problem:

Two different 12-V automobile batteries on a store shelf are rated at 600 and 850 “cold cranking amps.” Which has the smallest internal resistance?

Exercise:

Problem:

What are the advantages and disadvantages of connecting batteries in series? In parallel?

Exercise:

Problem:

Semitractor trucks use four large 12-V batteries. The starter system requires 24 V, while normal operation of the truck’s other electrical components utilizes 12 V. How could the four batteries be connected to produce 24 V? To produce 12 V? Why is 24 V better than 12 V for starting the truck’s engine (a very heavy load)?

Problem Exercises

Exercise:

Problem:

Standard automobile batteries have six lead-acid cells in series, creating a total emf of 12.0 V. What is the emf of an individual lead-acid cell?

Solution:

2.00 V

Exercise:

Problem:

Carbon-zinc dry cells (sometimes referred to as non-alkaline cells) have an emf of 1.54 V, and they are produced as single cells or in various combinations to form other voltages. (a) How many 1.54-V cells are needed to make the common 9-V battery used in many small electronic devices? (b) What is the actual emf of the approximately 9-V battery? (c) Discuss how internal resistance in the series connection of cells will affect the terminal voltage of this approximately 9-V battery.

Exercise:

Problem:

What is the output voltage of a 3.0000-V lithium cell in a digital wristwatch that draws 0.300 mA, if the cell's internal resistance is 2.00 Ω?

Solution:

2.9994 V

Exercise:

Problem:

(a) What is the terminal voltage of a large 1.54-V carbon-zinc dry cell used in a physics lab to supply 2.00 A to a circuit, if the cell's internal resistance is $0.100\ \Omega$? (b) How much electrical power does the cell produce? (c) What power goes to its load?

Exercise:

Problem:

What is the internal resistance of an automobile battery that has an emf of 12.0 V and a terminal voltage of 15.0 V while a current of 8.00 A is charging it?

Solution:

$0.375\ \Omega$

Exercise:

Problem:

(a) Find the terminal voltage of a 12.0-V motorcycle battery having a $0.600\text{-}\Omega$ internal resistance, if it is being charged by a current of 10.0 A. (b) What is the output voltage of the battery charger?

Exercise:

Problem:

A car battery with a 12-V emf and an internal resistance of $0.050\ \Omega$ is being charged with a current of 60 A. Note that in this process the battery is being charged. (a) What is the potential difference across its terminals? (b) At what rate is thermal energy being dissipated in the battery? (c) At what rate is electric energy being converted to chemical energy? (d) What are the answers to (a) and (b) when the battery is used to supply 60 A to the starter motor?

Exercise:

Problem:

The hot resistance of a flashlight bulb is $2.30\,\Omega$, and it is run by a 1.58-V alkaline cell having a 0.100-Ω internal resistance. (a) What current flows? (b) Calculate the power supplied to the bulb using I^2R_{bulb}. (c) Is this power the same as calculated using $\frac{V^2}{R_{\text{bulb}}}$?

Solution:

(a) 0.658 A

(b) 0.997 W

(c) 0.997 W; yes

Exercise:

Problem:

The label on a portable radio recommends the use of rechargeable nickel-cadmium cells (nicads), although they have a 1.25-V emf while alkaline cells have a 1.58-V emf. The radio has a 3.20-Ω resistance. (a) Draw a circuit diagram of the radio and its batteries. Now, calculate the power delivered to the radio. (b) When using Nicad cells each having an internal resistance of $0.0400\ \Omega$. (c) When using alkaline cells each having an internal resistance of $0.200\ \Omega$. (d) Does this difference seem significant, considering that the radio's effective resistance is lowered when its volume is turned up?

Exercise:

Problem:

An automobile starter motor has an equivalent resistance of $0.0500\,\Omega$ and is supplied by a 12.0-V battery with a 0.0100-Ω internal resistance. (a) What is the current to the motor? (b) What voltage is applied to it? (c) What power is supplied to the motor? (d) Repeat these calculations for when the battery connections are corroded and add $0.0900\,\Omega$ to the circuit. (Significant problems are caused by even small amounts of unwanted resistance in low-voltage, high-current applications.)

Solution:

(a) 200 A

(b) 10.0 V

(c) 2.00 kW

(d) $0.1000\,\Omega$*;* 80.0 A, 4.0 V, 320 W

Exercise:

Problem:

A child's electronic toy is supplied by three 1.58-V alkaline cells having internal resistances of $0.0200\,\Omega$ in series with a 1.53-V carbon-zinc dry cell having a 0.100-Ω internal resistance. The load resistance is $10.0\,\Omega$. (a) Draw a circuit diagram of the toy and its batteries. (b) What current flows? (c) How much power is supplied to the load? (d) What is the internal resistance of the dry cell if it goes bad, resulting in only 0.500 W being supplied to the load?

Exercise:

Problem:

(a) What is the internal resistance of a voltage source if its terminal voltage drops by 2.00 V when the current supplied increases by 5.00 A? (b) Can the emf of the voltage source be found with the information supplied?

Solution:

(a) $0.400\ \Omega$

(b) No, there is only one independent equation, so only r can be found.

Exercise:

Problem:

A person with body resistance between his hands of $10.0\ \text{k}\Omega$ accidentally grasps the terminals of a 20.0-kV power supply. (Do NOT do this!) (a) Draw a circuit diagram to represent the situation. (b) If the internal resistance of the power supply is $2000\,\Omega$, what is the current through his body? (c) What is the power dissipated in his body? (d) If the power supply is to be made safe by increasing its internal resistance, what should the internal resistance be for the maximum current in this situation to be 1.00 mA or less? (e) Will this modification compromise the effectiveness of the power supply for driving low-resistance devices? Explain your reasoning.

Exercise:

Problem:

Electric fish generate current with biological cells called electroplaques, which are physiological emf devices. The electroplaques in the South American eel are arranged in 140 rows, each row stretching horizontally along the body and each containing 5000 electroplaques. Each electroplaque has an emf of 0.15 V and internal resistance of $0.25\,\Omega$. If the water surrounding the fish has resistance of $800\,\Omega$, how much current can the eel produce in water from near its head to near its tail?

Exercise:

Problem: Integrated Concepts

A 12.0-V emf automobile battery has a terminal voltage of 16.0 V when being charged by a current of 10.0 A. (a) What is the battery's internal resistance? (b) What power is dissipated inside the battery? (c) At what rate (in $^{\circ}\mathrm{C}/\mathrm{min}$) will its temperature increase if its mass is 20.0 kg and it has a specific heat of $0.300\ \mathrm{kcal}/\mathrm{kg}\cdot{}^{\circ}\mathrm{C}$, assuming no heat escapes?

Exercise:

Problem: Unreasonable Results

A 1.58-V alkaline cell with a $0.200\text{-}\Omega$ internal resistance is supplying 8.50 A to a load. (a) What is its terminal voltage? (b) What is the value of the load resistance? (c) What is unreasonable about these results? (d) Which assumptions are unreasonable or inconsistent?

Solution:

(a) –0.120 V

(b) $-1.41\times10^{-2}\ \Omega$

(c) Negative terminal voltage; negative load resistance.

(d) The assumption that such a cell could provide 8.50 A is inconsistent with its internal resistance.

Exercise:

Problem: Unreasonable Results

(a) What is the internal resistance of a 1.54-V dry cell that supplies 1.00 W of power to a $15.0\text{-}\Omega$ bulb? (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Glossary

electromotive force (emf)
: the potential difference of a source of electricity when no current is flowing; measured in volts

internal resistance
: the amount of resistance within the voltage source

potential difference
: the difference in electric potential between two points in an electric circuit, measured in volts

terminal voltage
: the voltage measured across the terminals of a source of potential difference

Current

- Define electric current, ampere, and drift velocity
- Describe the direction of charge flow in conventional current.
- Use drift velocity to calculate current and vice versa.

Electric Current

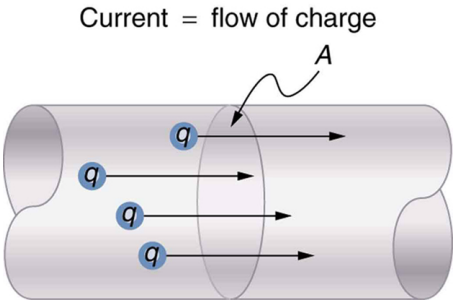
Electric current is defined to be the rate at which charge flows. A large current, such as that used to start a truck engine, moves a large amount of charge in a small time, whereas a small current, such as that used to operate a hand-held calculator, moves a small amount of charge over a long period of time. In equation form, **electric current** I is defined to be
Equation:

$$I = \frac{\Delta Q}{\Delta t},$$

where ΔQ is the amount of charge passing through a given area in time Δt. (As in previous chapters, initial time is often taken to be zero, in which case $\Delta t = t$.) (See [link].) The SI unit for current is the **ampere** (A), named for the French physicist André-Marie Ampère (1775–1836). Since $I = \Delta Q/\Delta t$, we see that an ampere is one coulomb per second:
Equation:

$$1 \text{ A} = 1 \text{ C/s}$$

Not only are fuses and circuit breakers rated in amperes (or amps), so are many electrical appliances.



The rate of flow of charge is current. An ampere is the flow of one coulomb through an area in one second.

Example:
Calculating Currents: Current in a Truck Battery and a Handheld Calculator
(a) What is the current involved when a truck battery sets in motion 720 C of charge in 4.00 s while starting an engine? (b) How long does it take 1.00 C of charge to flow through a handheld calculator if a 0.300-mA current is flowing?
Strategy
We can use the definition of current in the equation $I = \Delta Q/\Delta t$ to find the current in part (a), since charge and time are given. In part (b), we rearrange the definition of current and use the given values of charge and current to find the time required.
Solution for (a)
Entering the given values for charge and time into the definition of current gives
Equation:

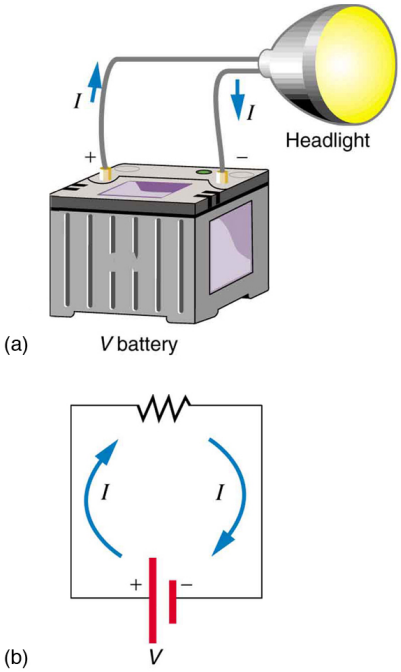
$$\begin{aligned} I &= \frac{\Delta Q}{\Delta t} = \frac{720\ \text{C}}{4.00\ \text{s}} = 180\ \text{C/s} \\ &= 180\ \text{A}. \end{aligned}$$

Discussion for (a)
This large value for current illustrates the fact that a large charge is moved in a small amount of time. The currents in these “starter motors” are fairly large because large frictional forces need to be overcome when setting something in motion.
Solution for (b)
Solving the relationship $I = \Delta Q/\Delta t$ for time Δt, and entering the known values for charge and current gives
Equation:

$$\begin{aligned} \Delta t &= \frac{\Delta Q}{I} = \frac{1.00\text{ C}}{0.300\times10^{-3}\text{ C/s}} \\ &= 3.33\times10^{3}\text{ s}. \end{aligned}$$

Discussion for (b)
This time is slightly less than an hour. The small current used by the hand-held calculator takes a much longer time to move a smaller charge than the large current of the truck starter. So why can we operate our calculators only seconds after turning them on? It’s because calculators require very little energy. Such small current and energy demands allow handheld calculators to operate from solar cells or to get many hours of use out of small batteries. Remember, calculators do not have moving parts in the same way that a truck engine has with cylinders and pistons, so the technology requires smaller currents.

[link] shows a simple circuit and the standard schematic representation of a battery, conducting path, and load (a resistor). Schematics are very useful in visualizing the main features of a circuit. A single schematic can represent a wide variety of situations. The schematic in [link] (b), for example, can represent anything from a truck battery connected to a headlight lighting the street in front of the truck to a small battery connected to a penlight lighting a keyhole in a door. Such schematics are useful because the analysis is the same for a wide variety of situations. We need to understand a few schematics to apply the concepts and analysis to many more situations.



(a) A simple electric circuit. A closed path for current to flow through is supplied by conducting wires connecting a load to the terminals of a battery. (b) In this schematic, the battery is represented by the two parallel red lines, conducting wires are shown as straight lines, and the zigzag represents the load. The schematic represents a wide

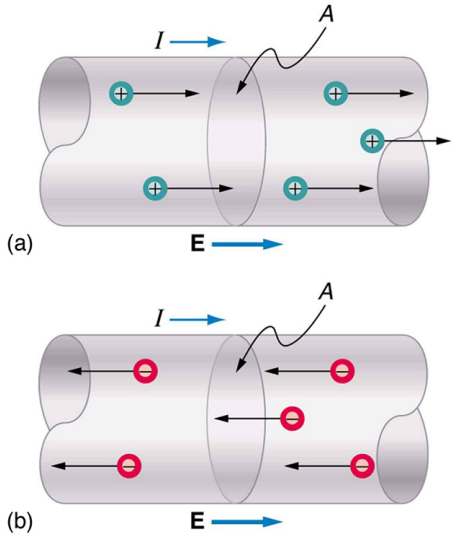
variety of similar
circuits.

Note that the direction of current flow in [link] is from positive to negative. *The direction of conventional current is the direction that positive charge would flow*. Depending on the situation, positive charges, negative charges, or both may move. In metal wires, for example, current is carried by electrons—that is, negative charges move. In ionic solutions, such as salt water, both positive and negative charges move. This is also true in nerve cells. A Van de Graaff generator used for nuclear research can produce a current of pure positive charges, such as protons. [link] illustrates the movement of charged particles that compose a current. The fact that conventional current is taken to be in the direction that positive charge would flow can be traced back to American politician and scientist Benjamin Franklin in the 1700s. He named the type of charge associated with electrons negative, long before they were known to carry current in so many situations. Franklin, in fact, was totally unaware of the small-scale structure of electricity.

It is important to realize that there is an electric field in conductors responsible for producing the current, as illustrated in [link]. Unlike static electricity, where a conductor in equilibrium cannot have an electric field in it, conductors carrying a current have an electric field and are not in static equilibrium. An electric field is needed to supply energy to move the charges.

Note:
Making Connections: Take-Home Investigation—Electric Current Illustration
Find a straw and little peas that can move freely in the straw. Place the straw flat on a table and fill the straw with peas. When you pop one pea in at one end, a different pea should pop out the other end. This demonstration is an analogy for an electric current. Identify what compares

to the electrons and what compares to the supply of energy. What other analogies can you find for an electric current?
Note that the flow of peas is based on the peas physically bumping into each other; electrons flow due to mutually repulsive electrostatic forces.



Current I is the rate at which charge moves through an area A, such as the cross-section of a wire. Conventional current is defined to move in the direction of the electric field. (a) Positive charges move in the direction of the electric field and the same direction as conventional current. (b) Negative charges move in the direction opposite to the electric field. Conventional

current is in the direction opposite to the movement of negative charge. The flow of electrons is sometimes referred to as electronic flow.

Example:
Calculating the Number of Electrons that Move through a Calculator
If the 0.300-mA current through the calculator mentioned in the [link] example is carried by electrons, how many electrons per second pass through it?
Strategy
The current calculated in the previous example was defined for the flow of positive charge. For electrons, the magnitude is the same, but the sign is opposite, $I_{\text{electrons}} = -0.300 \times 10^{-3}\ \text{C/s}$ .Since each electron (e^-) has a charge of $-1.60 \times 10^{-19}\ \text{C}$, we can convert the current in coulombs per second to electrons per second.
Solution
Starting with the definition of current, we have
Equation:

$$I_{\text{electrons}} = \frac{\Delta Q_{\text{electrons}}}{\Delta t} = \frac{-0.300 \times 10^{-3}\ \text{C}}{\text{s}}.$$

We divide this by the charge per electron, so that
Equation:

$$\begin{aligned} \frac{e^-}{\text{s}} &= \frac{-0.300 \times 10^{-3}\ \text{C}}{\text{s}} \times \frac{1\ e^-}{-1.60 \times 10^{-19}\ \text{C}} \\ &= 1.88 \times 10^{15}\ \frac{e^-}{\text{s}}. \end{aligned}$$

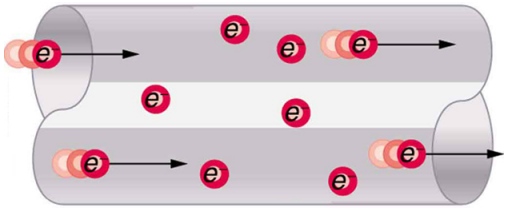
Discussion

There are so many charged particles moving, even in small currents, that individual charges are not noticed, just as individual water molecules are not noticed in water flow. Even more amazing is that they do not always keep moving forward like soldiers in a parade. Rather they are like a crowd of people with movement in different directions but a general trend to move forward. There are lots of collisions with atoms in the metal wire and, of course, with other electrons.

Drift Velocity

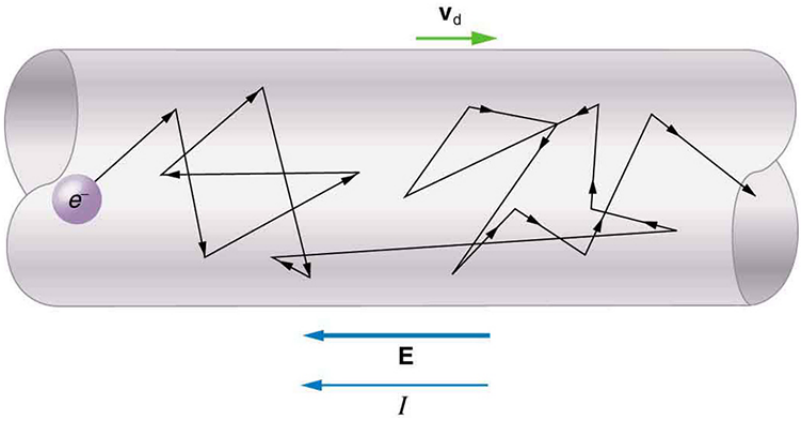
Electrical signals are known to move very rapidly. Telephone conversations carried by currents in wires cover large distances without noticeable delays. Lights come on as soon as a switch is flicked. Most electrical signals carried by currents travel at speeds on the order of $10^8\ \mathrm{m/s}$, a significant fraction of the speed of light. Interestingly, the individual charges that make up the current move *much* more slowly on average, typically drifting at speeds on the order of $10^{-4}\ \mathrm{m/s}$. How do we reconcile these two speeds, and what does it tell us about standard conductors?

The high speed of electrical signals results from the fact that the force between charges acts rapidly at a distance. Thus, when a free charge is forced into a wire, as in [link], the incoming charge pushes other charges ahead of it, which in turn push on charges farther down the line. The density of charge in a system cannot easily be increased, and so the signal is passed on rapidly. The resulting electrical shock wave moves through the system at nearly the speed of light. To be precise, this rapidly moving signal or shock wave is a rapidly propagating change in electric field.



When charged particles are forced into this volume of a conductor, an equal number are quickly forced to leave. The repulsion between like charges makes it difficult to increase the number of charges in a volume. Thus, as one charge enters, another leaves almost immediately, carrying the signal rapidly forward.

Good conductors have large numbers of free charges in them. In metals, the free charges are free electrons. [link] shows how free electrons move through an ordinary conductor. The distance that an individual electron can move between collisions with atoms or other electrons is quite small. The electron paths thus appear nearly random, like the motion of atoms in a gas. But there is an electric field in the conductor that causes the electrons to drift in the direction shown (opposite to the field, since they are negative). The **drift velocity** v_{d} is the average velocity of the free charges. Drift velocity is quite small, since there are so many free charges. If we have an estimate of the density of free electrons in a conductor, we can calculate the drift velocity for a given current. The larger the density, the lower the velocity required for a given current.



Free electrons moving in a conductor make many collisions with other electrons and atoms. The path of one electron is shown. The average velocity of the free charges is called the drift velocity, v_d, and it is in the direction opposite to the electric field for electrons. The collisions normally transfer energy to the conductor, requiring a constant supply of energy to maintain a steady current.

Note:
Conduction of Electricity and Heat
Good electrical conductors are often good heat conductors, too. This is because large numbers of free electrons can carry electrical current and can transport thermal energy.

The free-electron collisions transfer energy to the atoms of the conductor. The electric field does work in moving the electrons through a distance, but that work does not increase the kinetic energy (nor speed, therefore) of the electrons. The work is transferred to the conductor’s atoms, possibly

increasing temperature. Thus a continuous power input is required to keep a current flowing. An exception, of course, is found in superconductors, for reasons we shall explore in a later chapter. Superconductors can have a steady current without a continual supply of energy—a great energy savings. In contrast, the supply of energy can be useful, such as in a lightbulb filament. The supply of energy is necessary to increase the temperature of the tungsten filament, so that the filament glows.

Note:
Making Connections: Take-Home Investigation—Filament Observations
Find a lightbulb with a filament. Look carefully at the filament and describe its structure. To what points is the filament connected?

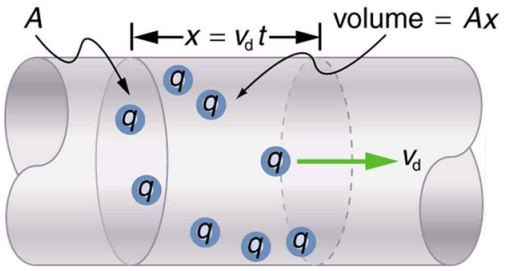
We can obtain an expression for the relationship between current and drift velocity by considering the number of free charges in a segment of wire, as illustrated in [link]. *The number of free charges per unit volume* is given the symbol n and depends on the material. The shaded segment has a volume Ax, so that the number of free charges in it is nAx. The charge ΔQ in this segment is thus qnAx, where q is the amount of charge on each carrier. (Recall that for electrons, q is -1.60×10^{-19} C.) Current is charge moved per unit time; thus, if all the original charges move out of this segment in time Δt, the current is
Equation:

$$I = \frac{\Delta Q}{\Delta t} = \frac{\mathrm{qnAx}}{\Delta t}.$$

Note that $x/\Delta t$ is the magnitude of the drift velocity, v_{d}, since the charges move an average distance x in a time Δt. Rearranging terms gives
Equation:

$$I = \mathrm{nqAv_d},$$

where I is the current through a wire of cross-sectional area A made of a material with a free charge density n. The carriers of the current each have charge q and move with a drift velocity of magnitude v_{d}.



All the charges in the shaded volume of this wire move out in a time t, having a drift velocity of magnitude $v_{\mathrm{d}} = x/t$. See text for further discussion.

Note that simple drift velocity is not the entire story. The speed of an electron is much greater than its drift velocity. In addition, not all of the electrons in a conductor can move freely, and those that do might move somewhat faster or slower than the drift velocity. So what do we mean by free electrons? Atoms in a metallic conductor are packed in the form of a lattice structure. Some electrons are far enough away from the atomic nuclei that they do not experience the attraction of the nuclei as much as the inner electrons do. These are the free electrons. They are not bound to a single atom but can instead move freely among the atoms in a “sea” of electrons. These free electrons respond by accelerating when an electric field is applied. Of course as they move they collide with the atoms in the lattice and other electrons, generating thermal energy, and the conductor gets warmer. In an insulator, the organization of the atoms and the structure do not allow for such free electrons.

Example:

Calculating Drift Velocity in a Common Wire

Calculate the drift velocity of electrons in a 12-gauge copper wire (which has a diameter of 2.053 mm) carrying a 20.0-A current, given that there is one free electron per copper atom. (Household wiring often contains 12-gauge copper wire, and the maximum current allowed in such wire is usually 20 A.) The density of copper is $8.80 \times 10^3 \text{ kg/m}^3$.

Strategy

We can calculate the drift velocity using the equation $I = nqAv_\text{d}$. The current $I = 20.0$ A is given, and $q = -1.60 \times 10^{-19}\text{C}$ is the charge of an electron. We can calculate the area of a cross-section of the wire using the formula $A = \pi r^2$, where r is one-half the given diameter, 2.053 mm. We are given the density of copper, $8.80 \times 10^3 \text{ kg/m}^3$, and the periodic table shows that the atomic mass of copper is 63.54 g/mol. We can use these two quantities along with Avogadro's number, $6.02 \times 10^{23} \text{ atoms/mol}$, to determine n, the number of free electrons per cubic meter.

Solution

First, calculate the density of free electrons in copper. There is one free electron per copper atom. Therefore, is the same as the number of copper atoms per m^3. We can now find n as follows:

Equation:

$$\begin{aligned} n &= \frac{1\ e^-}{\text{atom}} \times \frac{6.02 \times 10^{23}\ \text{atoms}}{\text{mol}} \times \frac{1\ \text{mol}}{63.54\ \text{g}} \times \frac{1000\ \text{g}}{\text{kg}} \times \frac{8.80 \times 10^3\ \text{kg}}{1\ \text{m}^3} \\ &= 8.342 \times 10^{28}\ e^-/\text{m}^3. \end{aligned}$$

The cross-sectional area of the wire is

Equation:

$$\begin{aligned} A &= \pi r^2 \\ &= \pi \left(\frac{2.053 \times 10^{-3}\ \text{m}}{2} \right)^2 \\ &= 3.310 \times 10^{-6}\ \text{m}^2. \end{aligned}$$

Rearranging $I = nqAv_\text{d}$ to isolate drift velocity gives

Equation:

$$v_{\text{d}} = \frac{I}{nqA}$$
$$= \frac{20.0\ \text{A}}{(8.342 \times 10^{28}/\text{m}^3)(-1.60 \times 10^{-19}\ \text{C})(3.310 \times 10^{-6}\ \text{m}^2)}$$
$$= -4.53 \times 10^{-4}\ \text{m/s}.$$

Discussion
The minus sign indicates that the negative charges are moving in the direction opposite to conventional current. The small value for drift velocity (on the order of 10^{-4} m/s) confirms that the signal moves on the order of 10^{12} times faster (about 10^{8} m/s) than the charges that carry it.

Section Summary

- Electric current I is the rate at which charge flows, given by
  Equation:

$$I = \frac{\Delta Q}{\Delta t},$$

  where ΔQ is the amount of charge passing through an area in time Δt.
- The direction of conventional current is taken as the direction in which positive charge moves.
- The SI unit for current is the ampere (A), where $1\ \text{A} = 1\ \text{C/s}$.
- Current is the flow of free charges, such as electrons and ions.
- Drift velocity v_{d} is the average speed at which these charges move.
- Current I is proportional to drift velocity v_{d}, as expressed in the relationship $I = \text{nqAv}_{\text{d}}$. Here, I is the current through a wire of cross-sectional area A. The wire's material has a free-charge density n, and each carrier has charge q and a drift velocity v_{d}.
- Electrical signals travel at speeds about 10^{12} times greater than the drift velocity of free electrons.

Conceptual Questions

Exercise:

Problem:

Can a wire carry a current and still be neutral—that is, have a total charge of zero? Explain.

Exercise:

Problem:

Car batteries are rated in ampere-hours ($\mathrm{A \cdot h}$). To what physical quantity do ampere-hours correspond (voltage, charge, . . .), and what relationship do ampere-hours have to energy content?

Exercise:

Problem:

If two different wires having identical cross-sectional areas carry the same current, will the drift velocity be higher or lower in the better conductor? Explain in terms of the equation $v_{\mathrm{d}} = \frac{I}{\mathrm{nqA}}$, by considering how the density of charge carriers n relates to whether or not a material is a good conductor.

Exercise:

Problem:

Why are two conducting paths from a voltage source to an electrical device needed to operate the device?

Exercise:

Problem:

In cars, one battery terminal is connected to the metal body. How does this allow a single wire to supply current to electrical devices rather than two wires?

Exercise:

Problem:

Why isn’t a bird sitting on a high-voltage power line electrocuted? Contrast this with the situation in which a large bird hits two wires simultaneously with its wings.

Problems & Exercises

Exercise:

Problem:

What is the current in milliamperes produced by the solar cells of a pocket calculator through which 4.00 C of charge passes in 4.00 h?

Solution:

0.278 mA

Exercise:

Problem:

A total of 600 C of charge passes through a flashlight in 0.500 h. What is the average current?

Exercise:

Problem:

What is the current when a typical static charge of $0.250\ \mu\text{C}$ moves from your finger to a metal doorknob in $1.00\ \mu\text{s}$?

Solution:

0.250 A

Exercise:

Problem:

Find the current when 2.00 nC jumps between your comb and hair over a 0.500 - μs time interval.

Exercise:

Problem:

A large lightning bolt had a 20,000-A current and moved 30.0 C of charge. What was its duration?

Solution:

1.50ms

Exercise:

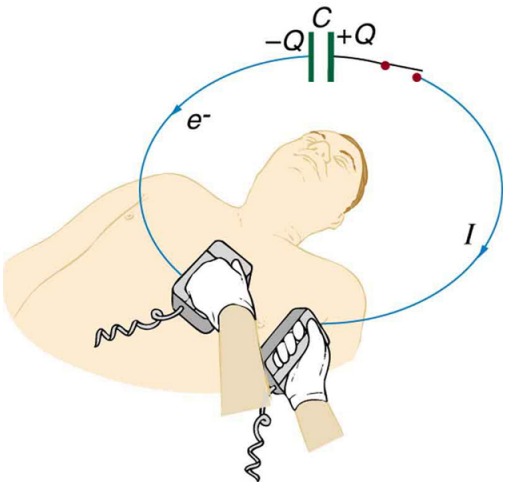
Problem:

The 200-A current through a spark plug moves 0.300 mC of charge. How long does the spark last?

Exercise:

Problem:

(a) A defibrillator sends a 6.00-A current through the chest of a patient by applying a 10,000-V potential as in the figure below. What is the resistance of the path? (b) The defibrillator paddles make contact with the patient through a conducting gel that greatly reduces the path resistance. Discuss the difficulties that would ensue if a larger voltage were used to produce the same current through the patient, but with the path having perhaps 50 times the resistance. (Hint: The current must be about the same, so a higher voltage would imply greater power. Use this equation for power: $P = I^2R$.)



The capacitor in a defibrillation unit drives a current through the heart of a patient.

Solution:

(a) $1.67\text{k}\Omega$

(b) If a 50 times larger resistance existed, keeping the current about the same, the power would be increased by a factor of about 50 (based on the equation $P = I^2R$), causing much more energy to be transferred to the skin, which could cause serious burns. The gel used reduces the resistance, and therefore reduces the power transferred to the skin.

Exercise:

Problem:

During open-heart surgery, a defibrillator can be used to bring a patient out of cardiac arrest. The resistance of the path is $500\ \Omega$ and a 10.0-mA current is needed. What voltage should be applied?

Exercise:

Problem:

(a) A defibrillator passes 12.0 A of current through the torso of a person for 0.0100 s. How much charge moves? (b) How many electrons pass through the wires connected to the patient? (See figure two problems earlier.)

Solution:

(a) 0.120 C

(b) $7.50 \times 10^{17} \text{electrons}$

Exercise:

Problem:

A clock battery wears out after moving 10,000 C of charge through the clock at a rate of 0.500 mA. (a) How long did the clock run? (b) How many electrons per second flowed?

Exercise:

Problem:

The batteries of a submerged non-nuclear submarine supply 1000 A at full speed ahead. How long does it take to move Avogadro's number (6.02×10^{23}) of electrons at this rate?

Solution:

96.3 s

Exercise:

Problem:

Electron guns are used in X-ray tubes. The electrons are accelerated through a relatively large voltage and directed onto a metal target, producing X-rays. (a) How many electrons per second strike the target if the current is 0.500 mA? (b) What charge strikes the target in 0.750 s?

Exercise:

Problem:

A large cyclotron directs a beam of He^{++} nuclei onto a target with a beam current of 0.250 mA. (a) How many He^{++} nuclei per second is this? (b) How long does it take for 1.00 C to strike the target? (c) How long before 1.00 mol of He^{++} nuclei strike the target?

Solution:

(a) $7.81 \times 10^{14}\ \mathrm{He}^{++}\ \mathrm{nuclei/s}$

(b) $4.00 \times 10^{3}\ \mathrm{s}$

(c) $7.71 \times 10^{8}\ \mathrm{s}$

Exercise:

Problem:

Repeat the above example on [link], but for a wire made of silver and given there is one free electron per silver atom.

Exercise:

Problem:

Using the results of the above example on [link], find the drift velocity in a copper wire of twice the diameter and carrying 20.0 A.

Solution:

$-1.13\times10^{-4}\text{m/s}$

Exercise:

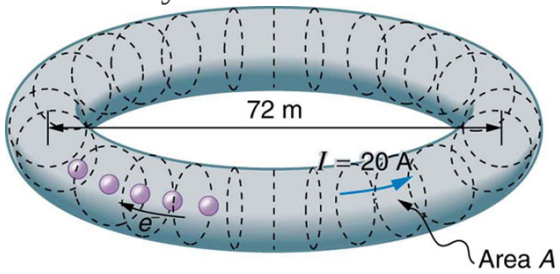
Problem:

A 14-gauge copper wire has a diameter of 1.628 mm. What magnitude current flows when the drift velocity is 1.00 mm/s? (See above example on [link] for useful information.)

Exercise:

Problem:

SPEAR, a storage ring about 72.0 m in diameter at the Stanford Linear Accelerator (closed in 2009), has a 20.0-A circulating beam of electrons that are moving at nearly the speed of light. (See [link].) How many electrons are in the beam?



Electrons circulating in the storage ring called SPEAR constitute a 20.0-A current. Because they travel close to the speed of light, each electron completes many orbits in each second.

Solution:

$9.42\times10^{13}\text{electrons}$

Glossary

electric current
: the rate at which charge flows, $I = \Delta Q/\Delta t$

ampere
: (amp) the SI unit for current; 1 A = 1 C/s

drift velocity
: the average velocity at which free charges flow in response to an electric field

Ohm’s Law: Resistance and Simple Circuits

- Explain the origin of Ohm’s law.
- Calculate voltages, currents, or resistances with Ohm’s law.
- Explain what an ohmic material is.
- Describe a simple circuit.

What drives current? We can think of various devices—such as batteries, generators, wall outlets, and so on—which are necessary to maintain a current. All such devices create a potential difference and are loosely referred to as voltage sources. When a voltage source is connected to a conductor, it applies a potential difference V that creates an electric field. The electric field in turn exerts force on charges, causing current.

Ohm’s Law

The current that flows through most substances is directly proportional to the voltage V applied to it. The German physicist Georg Simon Ohm (1787–1854) was the first to demonstrate experimentally that the current in a metal wire is *directly proportional to the voltage applied*:
Equation:

$$I \propto V.$$

This important relationship is known as **Ohm’s law**. It can be viewed as a cause-and-effect relationship, with voltage the cause and current the effect. This is an empirical law like that for friction—an experimentally observed phenomenon. Such a linear relationship doesn’t always occur.

Resistance and Simple Circuits

If voltage drives current, what impedes it? The electric property that impedes current (crudely similar to friction and air resistance) is called **resistance** R. Collisions of moving charges with atoms and molecules in a substance transfer energy to the substance and limit current. Resistance is defined as inversely proportional to current, or

Equation:

$$I \propto \frac{1}{R}.$$

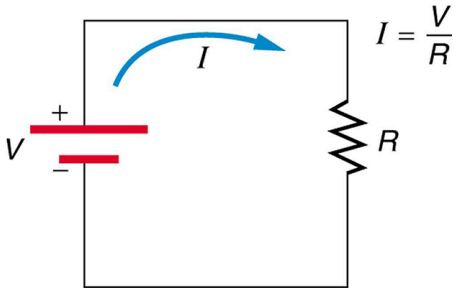
Thus, for example, current is cut in half if resistance doubles. Combining the relationships of current to voltage and current to resistance gives
Equation:

$$I = \frac{V}{R}.$$

This relationship is also called Ohm's law. Ohm's law in this form really defines resistance for certain materials. Ohm's law (like Hooke's law) is not universally valid. The many substances for which Ohm's law holds are called **ohmic**. These include good conductors like copper and aluminum, and some poor conductors under certain circumstances. Ohmic materials have a resistance R that is independent of voltage V and current I. An object that has simple resistance is called a *resistor*, even if its resistance is small. The unit for resistance is an **ohm** and is given the symbol Ω (upper case Greek omega). Rearranging $I = \mathrm{V}/\mathrm{R}$ gives $R = \mathrm{V}/\mathrm{I}$, and so the units of resistance are 1 ohm = 1 volt per ampere:
Equation:

$$1\ \Omega = 1\frac{V}{A}.$$

[link] shows the schematic for a simple circuit. A **simple circuit** has a single voltage source and a single resistor. The wires connecting the voltage source to the resistor can be assumed to have negligible resistance, or their resistance can be included in R.



A simple electric circuit in which a closed path for current to flow is supplied by conductors (usually metal wires) connecting a load to the terminals of a battery, represented by the red parallel lines. The zigzag symbol represents the single resistor and includes any resistance in the connections to the voltage source.

Example:
Calculating Resistance: An Automobile Headlight
What is the resistance of an automobile headlight through which 2.50 A flows when 12.0 V is applied to it?
Strategy
We can rearrange Ohm's law as stated by $I = \mathrm{V}/\mathrm{R}$ and use it to find the resistance.
Solution
Rearranging $I = \mathrm{V}/\mathrm{R}$ and substituting known values gives

Equation:

$$R = \frac{V}{I} = \frac{12.0\ \text{V}}{2.50\ \text{A}} = 4.80\ \Omega.$$

Discussion
This is a relatively small resistance, but it is larger than the cold resistance of the headlight. As we shall see in Resistance and Resistivity, resistance usually increases with temperature, and so the bulb has a lower resistance when it is first switched on and will draw considerably more current during its brief warm-up period.

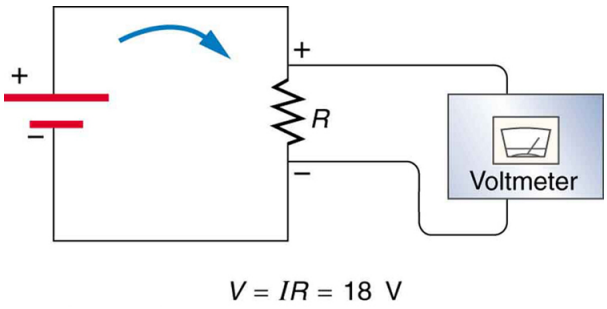
Resistances range over many orders of magnitude. Some ceramic insulators, such as those used to support power lines, have resistances of $10^{12}\ \Omega$ or more. A dry person may have a hand-to-foot resistance of $10^{5}\ \Omega$, whereas the resistance of the human heart is about $10^{3}\ \Omega$. A meter-long piece of large-diameter copper wire may have a resistance of $10^{-5}\ \Omega$, and superconductors have no resistance at all (they are non-ohmic). Resistance is related to the shape of an object and the material of which it is composed, as will be seen in Resistance and Resistivity.

Additional insight is gained by solving $I = \text{V}/\text{R}$ for V, yielding
Equation:

$$V = \text{IR}.$$

This expression for V can be interpreted as the *voltage drop across a resistor produced by the flow of current* I. The phrase IR *drop* is often used for this voltage. For instance, the headlight in [link] has an IR drop of 12.0 V. If voltage is measured at various points in a circuit, it will be seen to increase at the voltage source and decrease at the resistor. Voltage is similar to fluid pressure. The voltage source is like a pump, creating a pressure difference, causing current—the flow of charge. The resistor is like a pipe that reduces pressure and limits flow because of its resistance. Conservation of energy has important consequences here. The voltage source supplies

energy (causing an electric field and a current), and the resistor converts it to another form (such as thermal energy). In a simple circuit (one with a single simple resistor), the voltage supplied by the source equals the voltage drop across the resistor, since $\mathrm{PE} = q\Delta V$, and the same q flows through each. Thus the energy supplied by the voltage source and the energy converted by the resistor are equal. (See [link].)



The voltage drop across a resistor in a simple circuit equals the voltage output of the battery.

Note:
Making Connections: Conservation of Energy
In a simple electrical circuit, the sole resistor converts energy supplied by the source into another form. Conservation of energy is evidenced here by the fact that all of the energy supplied by the source is converted to another form by the resistor alone. We will find that conservation of energy has other important applications in circuits and is a powerful tool in circuit analysis.

Note:

PhET Explorations: Ohm's Law
See how the equation form of Ohm's law relates to a simple circuit. Adjust the voltage and resistance, and see the current change according to Ohm's law. The sizes of the symbols in the equation change to match the circuit diagram.
https://phet.colorado.edu/sims/html/ohms-law/latest/ohms-law_en.html

Section Summary

- A simple circuit *is* one in which there is a single voltage source and a single resistance.
- One statement of Ohm’s law gives the relationship between current I, voltage V, and resistance R in a simple circuit to be $I = \frac{V}{R}$.
- Resistance has units of ohms (Ω), related to volts and amperes by $1\ \Omega = 1\ \text{V}/\text{A}$.
- There is a voltage or IR drop across a resistor, caused by the current flowing through it, given by $V = \text{IR}$.

Conceptual Questions

Exercise:

Problem:

The IR drop across a resistor means that there is a change in potential or voltage across the resistor. Is there any change in current as it passes through a resistor? Explain.

Exercise:

Problem:

How is the IR drop in a resistor similar to the pressure drop in a fluid flowing through a pipe?

Problems & Exercises

Exercise:

Problem:

What current flows through the bulb of a 3.00-V flashlight when its hot resistance is $3.60\ \Omega$?

Solution:

0.833 A

Exercise:

Problem:

Calculate the effective resistance of a pocket calculator that has a 1.35-V battery and through which 0.200 mA flows.

Exercise:

Problem:

What is the effective resistance of a car’s starter motor when 150 A flows through it as the car battery applies 11.0 V to the motor?

Solution:

$7.33\times10^{-2}\ \Omega$

Exercise:

Problem:

How many volts are supplied to operate an indicator light on a DVD player that has a resistance of 140 Ω, given that 25.0 mA passes through it?

Exercise:

Problem:

(a) Find the voltage drop in an extension cord having a $0.0600\text{-}\Omega$ resistance and through which 5.00 A is flowing. (b) A cheaper cord utilizes thinner wire and has a resistance of $0.300\ \Omega$. What is the voltage drop in it when 5.00 A flows? (c) Why is the voltage to whatever appliance is being used reduced by this amount? What is the effect on the appliance?

Solution:

(a) 0.300 V

(b) 1.50 V

(c) The voltage supplied to whatever appliance is being used is reduced because the total voltage drop from the wall to the final output of the appliance is fixed. Thus, if the voltage drop across the extension cord is large, the voltage drop across the appliance is significantly decreased, so the power output by the appliance can be significantly decreased, reducing the ability of the appliance to work properly.

Exercise:

Problem:

A power transmission line is hung from metal towers with glass insulators having a resistance of $1.00\times10^{9}\ \Omega$. What current flows through the insulator if the voltage is 200 kV? (Some high-voltage lines are DC.)

Glossary

Ohm’s law
: an empirical relation stating that the current *I* is proportional to the potential difference *V*, $\propto V$; it is often written as $I = V/R$, where *R* is the resistance

resistance
: the electric property that impedes current; for ohmic materials, it is the ratio of voltage to current, $R = V/I$

ohm
: the unit of resistance, given by $1\Omega = 1$ V/A

ohmic
: a type of a material for which Ohm's law is valid

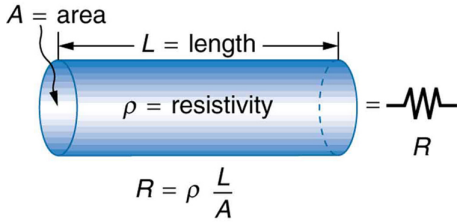
simple circuit
: a circuit with a single voltage source and a single resistor

Resistance and Resistivity

- Explain the concept of resistivity.
- Use resistivity to calculate the resistance of specified configurations of material.
- Use the thermal coefficient of resistivity to calculate the change of resistance with temperature.

Material and Shape Dependence of Resistance

The resistance of an object depends on its shape and the material of which it is composed. The cylindrical resistor in [link] is easy to analyze, and, by so doing, we can gain insight into the resistance of more complicated shapes. As you might expect, the cylinder's electric resistance R is directly proportional to its length L, similar to the resistance of a pipe to fluid flow. The longer the cylinder, the more collisions charges will make with its atoms. The greater the diameter of the cylinder, the more current it can carry (again similar to the flow of fluid through a pipe). In fact, R is inversely proportional to the cylinder's cross-sectional area A.



A uniform cylinder of length L and cross-sectional area A. Its resistance to the flow of current is similar to the resistance posed by a pipe to fluid flow. The longer the cylinder, the greater its

resistance. The larger its cross-sectional area A, the smaller its resistance.

For a given shape, the resistance depends on the material of which the object is composed. Different materials offer different resistance to the flow of charge. We define the **resistivity** ρ of a substance so that the **resistance** R of an object is directly proportional to ρ. Resistivity ρ is an *intrinsic* property of a material, independent of its shape or size. The resistance R of a uniform cylinder of length L, of cross-sectional area A, and made of a material with resistivity ρ, is

Equation:

$$R = \frac{\rho L}{A}.$$

[link] gives representative values of ρ. The materials listed in the table are separated into categories of conductors, semiconductors, and insulators, based on broad groupings of resistivities. Conductors have the smallest resistivities, and insulators have the largest; semiconductors have intermediate resistivities. Conductors have varying but large free charge densities, whereas most charges in insulators are bound to atoms and are not free to move. Semiconductors are intermediate, having far fewer free charges than conductors, but having properties that make the number of free charges depend strongly on the type and amount of impurities in the semiconductor. These unique properties of semiconductors are put to use in modern electronics, as will be explored in later chapters.

Material	Resistivity ρ ($\Omega \cdot \mathrm{m}$)
Conductors	
Silver	1.59×10^{-8}
Copper	1.72×10^{-8}
Gold	2.44×10^{-8}
Aluminum	2.65×10^{-8}
Tungsten	5.6×10^{-8}
Iron	9.71×10^{-8}
Platinum	10.6×10^{-8}
Steel	20×10^{-8}
Lead	22×10^{-8}

Material	Resistivity ρ ($\Omega \cdot \mathrm{m}$)
Manganin (Cu, Mn, Ni alloy)	44×10^{-8}
Constantan (Cu, Ni alloy)	49×10^{-8}
Mercury	96×10^{-8}
Nichrome (Ni, Fe, Cr alloy)	100×10^{-8}
Semiconductors[footnote] Values depend strongly on amounts and types of impurities	
Carbon (pure)	3.5×10^{-5}
Carbon	$(3.5 - 60) \times 10^{-5}$
Germanium (pure)	600×10^{-3}
Germanium	$(1 - 600) \times 10^{-3}$

Material	**Resistivity ρ ($\Omega \cdot \mathrm{m}$)**
Silicon (pure)	2300
Silicon	$0.1–2300$
Insulators	
Amber	5×10^{14}
Glass	$10^{9} - 10^{14}$
Lucite	$>10^{13}$
Mica	$10^{11} - 10^{15}$
Quartz (fused)	75×10^{16}
Rubber (hard)	$10^{13} - 10^{16}$
Sulfur	10^{15}

Material	Resistivity ρ ($\Omega \cdot \mathrm{m}$)
Teflon	$>10^{13}$
Wood	$10^{8}-10^{11}$

Resistivities ρ of Various materials at 20ºC

Example:
Calculating Resistor Diameter: A Headlight Filament
A car headlight filament is made of tungsten and has a cold resistance of $0.350\ \Omega$. If the filament is a cylinder 4.00 cm long (it may be coiled to save space), what is its diameter?
Strategy
We can rearrange the equation $R = \frac{\rho L}{A}$ to find the cross-sectional area A of the filament from the given information. Then its diameter can be found by assuming it has a circular cross-section.
Solution
The cross-sectional area, found by rearranging the expression for the resistance of a cylinder given in $R = \frac{\rho L}{A}$, is
Equation:

$$A = \frac{\rho L}{R}.$$

Substituting the given values, and taking ρ from [link], yields
Equation:

$$\begin{aligned} A &= \frac{(5.6\times10^{-8}\ \Omega\cdot\mathrm{m})(4.00\times10^{-2}\ \mathrm{m})}{0.350\ \Omega} \\ &= 6.40\times10^{-9}\ \mathrm{m}^2. \end{aligned}$$

The area of a circle is related to its diameter D by
Equation:

$$A = \frac{\pi D^2}{4}.$$

Solving for the diameter D, and substituting the value found for A, gives
Equation:

$$\begin{aligned} D &= 2\left(\frac{A}{p}\right)^{\frac{1}{2}} = 2\left(\frac{6.40\times10^{-9}\ \text{m}^2}{3.14}\right)^{\frac{1}{2}} \\ &= 9.0\times10^{-5}\ \text{m}. \end{aligned}$$

Discussion
The diameter is just under a tenth of a millimeter. It is quoted to only two digits, because ρ is known to only two digits.

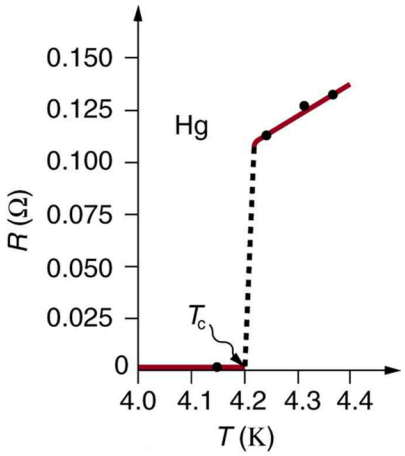
Temperature Variation of Resistance

The resistivity of all materials depends on temperature. Some even become superconductors (zero resistivity) at very low temperatures. (See [link].) Conversely, the resistivity of conductors increases with increasing temperature. Since the atoms vibrate more rapidly and over larger distances at higher temperatures, the electrons moving through a metal make more collisions, effectively making the resistivity higher. Over relatively small temperature changes (about 100ºC or less), resistivity ρ varies with temperature change ΔT as expressed in the following equation
Equation:

$$\rho = \rho_0(1 + \alpha\Delta T),$$

where ρ_0 is the original resistivity and α is the **temperature coefficient of resistivity**. (See the values of α in [link] below.) For larger temperature changes, α may vary or a nonlinear equation may be needed to find ρ. Note

that α is positive for metals, meaning their resistivity increases with temperature. Some alloys have been developed specifically to have a small temperature dependence. Manganin (which is made of copper, manganese and nickel), for example, has α close to zero (to three digits on the scale in [link]), and so its resistivity varies only slightly with temperature. This is useful for making a temperature-independent resistance standard, for example.



The resistance of a sample of mercury is zero at very low temperatures—it is a superconductor up to about 4.2 K. Above that critical temperature, its resistance makes a sudden jump and then increases nearly linearly with temperature.

Material	**Coefficient α(1/°C)[footnote] Values at 20°C.**
Conductors	
Silver	3.8×10^{-3}
Copper	3.9×10^{-3}
Gold	3.4×10^{-3}
Aluminum	3.9×10^{-3}
Tungsten	4.5×10^{-3}
Iron	5.0×10^{-3}
Platinum	3.93×10^{-3}
Lead	3.9×10^{-3}
Manganin (Cu, Mn, Ni alloy)	0.000×10^{-3}

Material	**Coefficient α(1/°C)[footnote] Values at 20°C.**
Constantan (Cu, Ni alloy)	0.002×10^{-3}
Mercury	0.89×10^{-3}
Nichrome (Ni, Fe, Cr alloy)	0.4×10^{-3}
Semiconductors	
Carbon (pure)	-0.5×10^{-3}
Germanium (pure)	-50×10^{-3}
Silicon (pure)	-70×10^{-3}

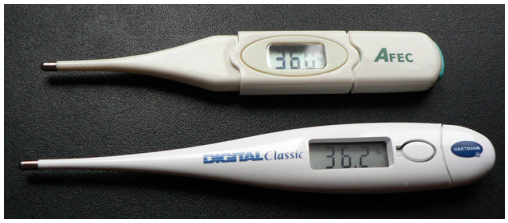
Tempature Coefficients of Resistivity α

Note also that α is negative for the semiconductors listed in [link], meaning that their resistivity decreases with increasing temperature. They become better conductors at higher temperature, because increased thermal agitation increases the number of free charges available to carry current. This property of decreasing ρ with temperature is also related to the type and amount of impurities present in the semiconductors.

The resistance of an object also depends on temperature, since R_0 is directly proportional to ρ. For a cylinder we know $R = \rho L/A$, and so, if L and A do not change greatly with temperature, R will have the same temperature dependence as ρ. (Examination of the coefficients of linear expansion shows them to be about two orders of magnitude less than typical temperature coefficients of resistivity, and so the effect of temperature on L and A is about two orders of magnitude less than on ρ.) Thus,
Equation:

$$R = R_0(1 + \alpha\Delta T)$$

is the temperature dependence of the resistance of an object, where R_0 is the original resistance and R is the resistance after a temperature change ΔT. Numerous thermometers are based on the effect of temperature on resistance. (See [link].) One of the most common is the thermistor, a semiconductor crystal with a strong temperature dependence, the resistance of which is measured to obtain its temperature. The device is small, so that it quickly comes into thermal equilibrium with the part of a person it touches.



These familiar thermometers are based on the automated measurement of a thermistor’s temperature-dependent resistance. (credit: Biol, Wikimedia Commons)

Example:

Calculating Resistance: Hot-Filament Resistance

Although caution must be used in applying $\rho = \rho_0(1 + \alpha\Delta T)$ and $R = R_0(1 + \alpha\Delta T)$ for temperature changes greater than 100ºC, for tungsten the equations work reasonably well for very large temperature changes. What, then, is the resistance of the tungsten filament in the previous example if its temperature is increased from room temperature (20ºC) to a typical operating temperature of 2850ºC?

Strategy

This is a straightforward application of $R = R_0(1 + \alpha\Delta T)$, since the original resistance of the filament was given to be $R_0 = 0.350\ \Omega$, and the temperature change is $\Delta T = 2830ºC$.

Solution

The hot resistance R is obtained by entering known values into the above equation:

Equation:

$$\begin{aligned} R &= R_0(1 + \alpha\Delta T) \\ &= (0.350\ \Omega)[1 + (4.5\times10^{-3}/ºC)(2830ºC)] \\ &= 4.8\ \Omega. \end{aligned}$$

Discussion

This value is consistent with the headlight resistance example in Ohm’s Law: Resistance and Simple Circuits.

Note:

PhET Explorations: Resistance in a Wire

Learn about the physics of resistance in a wire. Change its resistivity, length, and area to see how they affect the wire's resistance. The sizes of the symbols in the equation change along with the diagram of a wire.

https://phet.colorado.edu/sims/html/resistance-in-a-wire/latest/resistance-in-a-wire_en.html

Section Summary

- The resistance R of a cylinder of length L and cross-sectional area A is $R = \frac{\rho L}{A}$, where ρ is the resistivity of the material.
- Values of ρ in [link] show that materials fall into three groups—*conductors, semiconductors, and insulators*.
- Temperature affects resistivity; for relatively small temperature changes ΔT, resistivity is $\rho = \rho_0(1 + \alpha\Delta T)$, where ρ_0 is the original resistivity and α is the temperature coefficient of resistivity.
- [link] gives values for α, the temperature coefficient of resistivity.
- The resistance R of an object also varies with temperature: $R = R_0(1 + \alpha\Delta T)$, where R_0 is the original resistance, and R is the resistance after the temperature change.

Conceptual Questions

Exercise:

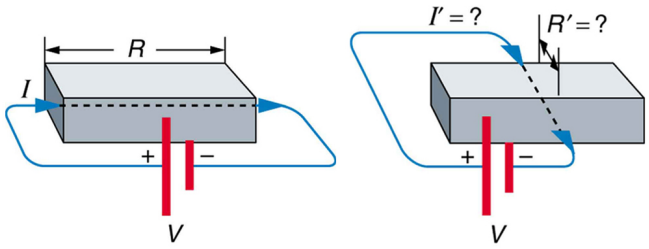
Problem:

In which of the three semiconducting materials listed in [link] do impurities supply free charges? (Hint: Examine the range of resistivity for each and determine whether the pure semiconductor has the higher or lower conductivity.)

Exercise:

Problem:

Does the resistance of an object depend on the path current takes through it? Consider, for example, a rectangular bar—is its resistance the same along its length as across its width? (See [link].)



Does current taking two different paths through the same object encounter different resistance?

Exercise:

Problem:

If aluminum and copper wires of the same length have the same resistance, which has the larger diameter? Why?

Exercise:

Problem:

Explain why $R = R_0(1 + \alpha\Delta T)$ for the temperature variation of the resistance R of an object is not as accurate as $\rho = \rho_0(1 + \alpha\Delta T)$, which gives the temperature variation of resistivity ρ.

Problems & Exercises

Exercise:

Problem:

What is the resistance of a 20.0-m-long piece of 12-gauge copper wire having a 2.053-mm diameter?

Solution:

0.104 Ω

Exercise:

Problem:

The diameter of 0-gauge copper wire is 8.252 mm. Find the resistance of a 1.00-km length of such wire used for power transmission.

Exercise:

Problem:

If the 0.100-mm diameter tungsten filament in a light bulb is to have a resistance of $0.200\ \Omega$ at $20.0^{\circ}\mathrm{C}$, how long should it be?

Solution:

$2.8\times10^{-2}\ \mathrm{m}$

Exercise:

Problem:

Find the ratio of the diameter of aluminum to copper wire, if they have the same resistance per unit length (as they might in household wiring).

Exercise:

Problem:

What current flows through a 2.54-cm-diameter rod of pure silicon that is 20.0 cm long, when $1.00\times10^{3}\ \mathrm{V}$ is applied to it? (Such a rod may be used to make nuclear-particle detectors, for example.)

Solution:

$1.10\times10^{-3}\ \mathrm{A}$

Exercise:

Problem:

(a) To what temperature must you raise a copper wire, originally at 20.0°C, to double its resistance, neglecting any changes in dimensions? (b) Does this happen in household wiring under ordinary circumstances?

Exercise:

Problem:

A resistor made of Nichrome wire is used in an application where its resistance cannot change more than 1.00% from its value at 20.0°C. Over what temperature range can it be used?

Solution:

-5°C to 45°C

Exercise:

Problem:

Of what material is a resistor made if its resistance is 40.0% greater at 100°C than at 20.0°C?

Exercise:

Problem:

An electronic device designed to operate at any temperature in the range from -10.0°C to 55.0°C contains pure carbon resistors. By what factor does their resistance increase over this range?

Solution:

1.03

Exercise:

Problem:

(a) Of what material is a wire made, if it is 25.0 m long with a 0.100 mm diameter and has a resistance of $77.7\ \Omega$ at $20.0^{\circ}\mathrm{C}$? (b) What is its resistance at $150^{\circ}\mathrm{C}$?

Exercise:

Problem:

Assuming a constant temperature coefficient of resistivity, what is the maximum percent decrease in the resistance of a constantan wire starting at $20.0^{\circ}\mathrm{C}$?

Solution:

0.06%

Exercise:

Problem:

A wire is drawn through a die, stretching it to four times its original length. By what factor does its resistance increase?

Exercise:

Problem:

A copper wire has a resistance of $0.500\ \Omega$ at $20.0^{\circ}\mathrm{C}$, and an iron wire has a resistance of $0.525\ \Omega$ at the same temperature. At what temperature are their resistances equal?

Solution:

$-17^{\circ}\mathrm{C}$

Exercise:

Problem:

(a) Digital medical thermometers determine temperature by measuring the resistance of a semiconductor device called a thermistor (which has $\alpha = -0.0600/^{\circ}\text{C}$) when it is at the same temperature as the patient. What is a patient's temperature if the thermistor's resistance at that temperature is 82.0% of its value at 37.0°C (normal body temperature)? (b) The negative value for α may not be maintained for very low temperatures. Discuss why and whether this is the case here. (Hint: Resistance can't become negative.)

Exercise:

Problem: Integrated Concepts

(a) Redo [link] taking into account the thermal expansion of the tungsten filament. You may assume a thermal expansion coefficient of $12 \times 10^{-6}/^{\circ}\text{C}$. (b) By what percentage does your answer differ from that in the example?

Solution:

(a) $4.7\ \Omega$ (total)

(b) 3.0% decrease

Exercise:

Problem: Unreasonable Results

(a) To what temperature must you raise a resistor made of constantan to double its resistance, assuming a constant temperature coefficient of resistivity? (b) To cut it in half? (c) What is unreasonable about these results? (d) Which assumptions are unreasonable, or which premises are inconsistent?

Glossary

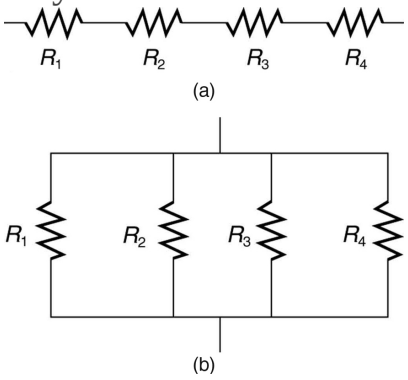
resistivity
: an intrinsic property of a material, independent of its shape or size, directly proportional to the resistance, denoted by ρ

temperature coefficient of resistivity
: an empirical quantity, denoted by α, which describes the change in resistance or resistivity of a material with temperature

Resistors in Series and Parallel

- Draw a circuit with resistors in parallel and in series.
- Calculate the voltage drop of a current across a resistor using Ohm’s law.
- Contrast the way total resistance is calculated for resistors in series and in parallel.
- Explain why total resistance of a parallel circuit is less than the smallest resistance of any of the resistors in that circuit.
- Calculate total resistance of a circuit that contains a mixture of resistors connected in series and in parallel.

Most circuits have more than one component, called a **resistor** that limits the flow of charge in the circuit. A measure of this limit on charge flow is called **resistance**. The simplest combinations of resistors are the series and parallel connections illustrated in [link]. The total resistance of a combination of resistors depends on both their individual values and how they are connected.

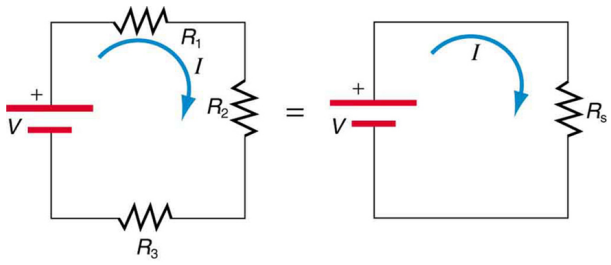


(a) A series connection of resistors. (b) A parallel connection of resistors.

Resistors in Series

When are resistors in **series**? Resistors are in series whenever the flow of charge, called the **current**, must flow through devices sequentially. For example, if current flows through a person holding a screwdriver and into the Earth, then R_1 in [link](a) could be the resistance of the screwdriver's shaft, R_2 the resistance of its handle, R_3 the person's body resistance, and R_4 the resistance of her shoes.

[link] shows resistors in series connected to a **voltage** source. It seems reasonable that the total resistance is the sum of the individual resistances, considering that the current has to pass through each resistor in sequence. (This fact would be an advantage to a person wishing to avoid an electrical shock, who could reduce the current by wearing high-resistance rubber-soled shoes. It could be a disadvantage if one of the resistances were a faulty high-resistance cord to an appliance that would reduce the operating current.)



Three resistors connected in series to a battery (left) and the equivalent single or series resistance (right).

To verify that resistances in series do indeed add, let us consider the loss of electrical power, called a **voltage drop**, in each resistor in [link].

According to **Ohm's law**, the voltage drop, V, across a resistor when a current flows through it is calculated using the equation $V = \text{IR}$, where I equals the current in amps (A) and R is the resistance in ohms (Ω). Another

way to think of this is that V is the voltage necessary to make a current I flow through a resistance R.

So the voltage drop across R_1 is $V_1 = IR_1$, that across R_2 is $V_2 = IR_2$, and that across R_3 is $V_3 = IR_3$. The sum of these voltages equals the voltage output of the source; that is,
Equation:

$$V = V_1 + V_2 + V_3.$$

This equation is based on the conservation of energy and conservation of charge. Electrical potential energy can be described by the equation $\mathrm{PE} = \mathrm{qV}$, where q is the electric charge and V is the voltage. Thus the energy supplied by the source is qV, while that dissipated by the resistors is
Equation:

$$qV_1 + qV_2 + qV_3.$$

Note:
Connections: Conservation Laws
The derivations of the expressions for series and parallel resistance are based on the laws of conservation of energy and conservation of charge, which state that total charge and total energy are constant in any process. These two laws are directly involved in all electrical phenomena and will be invoked repeatedly to explain both specific effects and the general behavior of electricity.

These energies must be equal, because there is no other source and no other destination for energy in the circuit. Thus, $qV = qV_1 + qV_2 + qV_3$. The charge q cancels, yielding $V = V_1 + V_2 + V_3$, as stated. (Note that the same amount of charge passes through the battery and each resistor in a given amount of time, since there is no capacitance to store charge, there is no place for charge to leak, and charge is conserved.)

Now substituting the values for the individual voltages gives
Equation:

$$V = IR_1 + IR_2 + IR_3 = I(R_1 + R_2 + R_3).$$

Note that for the equivalent single series resistance R_s, we have
Equation:

$$V = IR_\text{s}.$$

This implies that the total or equivalent series resistance R_s of three resistors is $R_\text{s} = R_1 + R_2 + R_3$.

This logic is valid in general for any number of resistors in series; thus, the total resistance R_s of a series connection is
Equation:

$$R_\text{s} = R_1 + R_2 + R_3 + ...,$$

as proposed. Since all of the current must pass through each resistor, it experiences the resistance of each, and resistances in series simply add up.

Example:
Calculating Resistance, Current, Voltage Drop, and Power Dissipation: Analysis of a Series Circuit
Suppose the voltage output of the battery in [link] is 12.0 V, and the resistances are $R_1 = 1.00\ \Omega$, $R_2 = 6.00\ \Omega$, and $R_3 = 13.0\ \Omega$. (a) What is the total resistance? (b) Find the current. (c) Calculate the voltage drop in each resistor, and show these add to equal the voltage output of the source. (d) Calculate the power dissipated by each resistor. (e) Find the power output of the source, and show that it equals the total power dissipated by the resistors.
Strategy and Solution for (a)

The total resistance is simply the sum of the individual resistances, as given by this equation:

Equation:

$$\begin{aligned} R_{\text{s}} &= R_1 + R_2 + R_3 \\ &= 1.00\ \Omega + 6.00\ \Omega + 13.0\ \Omega \\ &= 20.0\ \Omega. \end{aligned}$$

Strategy and Solution for (b)

The current is found using Ohm's law, $V = \text{IR}$. Entering the value of the applied voltage and the total resistance yields the current for the circuit:

Equation:

$$I = \frac{V}{R_{\text{s}}} = \frac{12.0\ \text{V}}{20.0\ \Omega} = 0.600\ \text{A}.$$

Strategy and Solution for (c)

The voltage—or IR drop—in a resistor is given by Ohm's law. Entering the current and the value of the first resistance yields

Equation:

$$V_1 = IR_1 = (0.600\ \text{A})(1.0\ \Omega) = 0.600\ \text{V}.$$

Similarly,

Equation:

$$V_2 = IR_2 = (0.600\ \text{A})(6.0\ \Omega) = 3.60\ \text{V}$$

and

Equation:

$$V_3 = IR_3 = (0.600\ \text{A})(13.0\ \Omega) = 7.80\ \text{V}.$$

Discussion for (c)

The three IR drops add to 12.0 V, as predicted:

Equation:

$$V_1 + V_2 + V_3 = (0.600 + 3.60 + 7.80)\ \text{V} = 12.0\ \text{V}.$$

Strategy and Solution for (d)
The easiest way to calculate power in watts (W) dissipated by a resistor in a DC circuit is to use **Joule's law**, $P = \text{IV}$, where P is electric power. In this case, each resistor has the same full current flowing through it. By substituting Ohm's law $V = \text{IR}$ into Joule's law, we get the power dissipated by the first resistor as
Equation:

$$P_1 = I^2R_1 = (0.600\ \text{A})^2(1.00\ \Omega) = 0.360\ \text{W}.$$

Similarly,
Equation:

$$P_2 = I^2R_2 = (0.600\ \text{A})^2(6.00\ \Omega) = 2.16\ \text{W}$$

and
Equation:

$$P_3 = I^2R_3 = (0.600\ \text{A})^2(13.0\ \Omega) = 4.68\ \text{W}.$$

Discussion for (d)
Power can also be calculated using either $P = \text{IV}$ or $P = \frac{V^2}{R}$, where V is the voltage drop across the resistor (not the full voltage of the source). The same values will be obtained.
Strategy and Solution for (e)
The easiest way to calculate power output of the source is to use $P = \text{IV}$, where V is the source voltage. This gives
Equation:

$$P = (0.600\ \text{A})(12.0\ \text{V}) = 7.20\ \text{W}.$$

Discussion for (e)
Note, coincidentally, that the total power dissipated by the resistors is also 7.20 W, the same as the power put out by the source. That is,
Equation:

$$P_1 + P_2 + P_3 = (0.360 + 2.16 + 4.68)\ \text{W} = 7.20\ \text{W}.$$

Power is energy per unit time (watts), and so conservation of energy requires the power output of the source to be equal to the total power dissipated by the resistors.

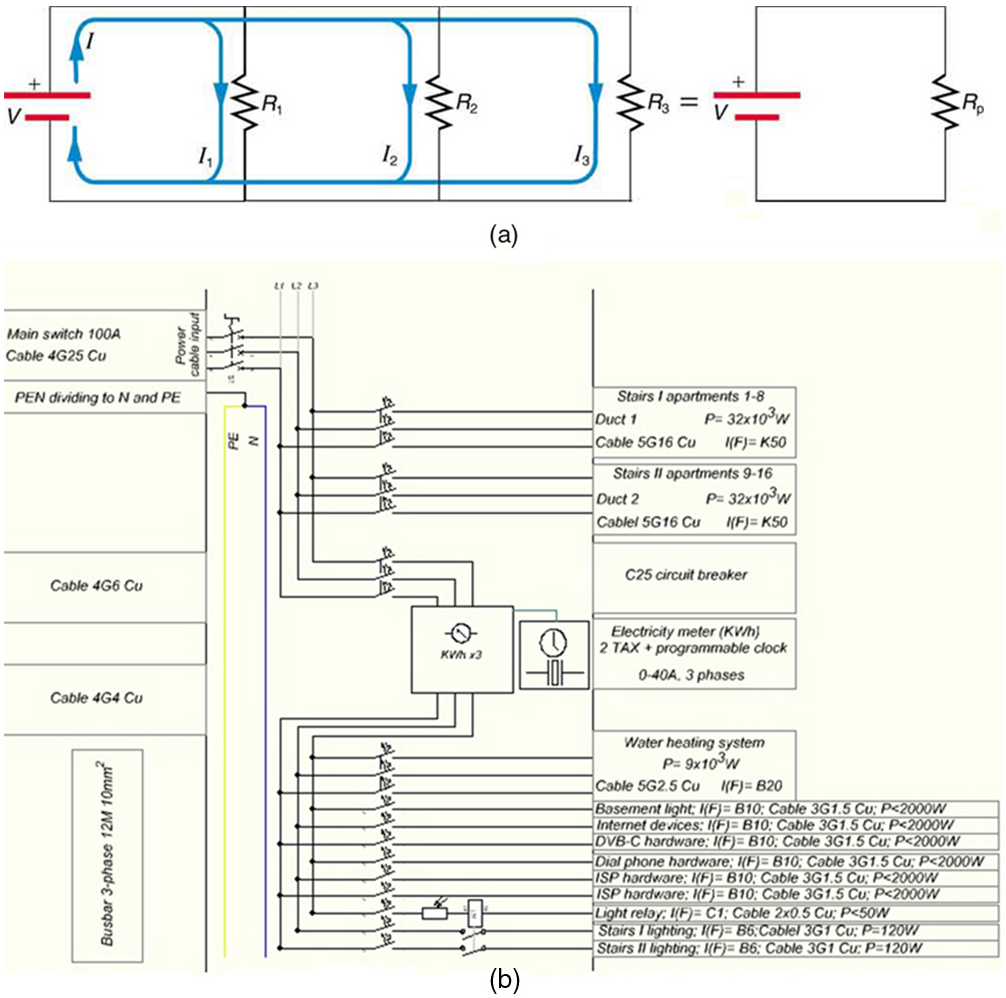
Note:
Major Features of Resistors in Series

1. Series resistances add: $R_s = R_1 + R_2 + R_3 + ...$
2. The same current flows through each resistor in series.
3. Individual resistors in series do not get the total source voltage, but divide it.

Resistors in Parallel

[link] shows resistors in **parallel**, wired to a voltage source. Resistors are in parallel when each resistor is connected directly to the voltage source by connecting wires having negligible resistance. Each resistor thus has the full voltage of the source applied to it.

Each resistor draws the same current it would if it alone were connected to the voltage source (provided the voltage source is not overloaded). For example, an automobile's headlights, radio, and so on, are wired in parallel, so that they utilize the full voltage of the source and can operate completely independently. The same is true in your house, or any building. (See [link] (b).)



(a) Three resistors connected in parallel to a battery and the equivalent single or parallel resistance. (b) Electrical power setup in a house. (credit: Dmitry G, Wikimedia Commons)

To find an expression for the equivalent parallel resistance R_{p}, let us consider the currents that flow and how they are related to resistance. Since each resistor in the circuit has the full voltage, the currents flowing through the individual resistors are $I_1 = \frac{V}{R_1}$, $I_2 = \frac{V}{R_2}$, and $I_3 = \frac{V}{R_3}$. Conservation of charge implies that the total current I produced by the source is the sum of these currents:

Equation:

$$I = I_1 + I_2 + I_3.$$

Substituting the expressions for the individual currents gives
Equation:

$$I = \frac{V}{R_1} + \frac{V}{R_2} + \frac{V}{R_3} = V\left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}\right).$$

Note that Ohm’s law for the equivalent single resistance gives
Equation:

$$I = \frac{V}{R_\text{p}} = V\left(\frac{1}{R_\text{p}}\right).$$

The terms inside the parentheses in the last two equations must be equal. Generalizing to any number of resistors, the total resistance R_p of a parallel connection is related to the individual resistances by
Equation:

$$\frac{1}{R_\text{p}} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_{.3}} + ...$$

This relationship results in a total resistance R_p that is less than the smallest of the individual resistances. (This is seen in the next example.) When resistors are connected in parallel, more current flows from the source than would flow for any of them individually, and so the total resistance is lower.

Example:
Calculating Resistance, Current, Power Dissipation, and Power Output: Analysis of a Parallel Circuit
Let the voltage output of the battery and resistances in the parallel connection in [link] be the same as the previously considered series

connection: $V = 12.0$ V, $R_1 = 1.00\ \Omega$, $R_2 = 6.00\ \Omega$, and $R_3 = 13.0\ \Omega$. (a) What is the total resistance? (b) Find the total current. (c) Calculate the currents in each resistor, and show these add to equal the total current output of the source. (d) Calculate the power dissipated by each resistor. (e) Find the power output of the source, and show that it equals the total power dissipated by the resistors.

Strategy and Solution for (a)

The total resistance for a parallel combination of resistors is found using the equation below. Entering known values gives

Equation:

$$\frac{1}{R_{\text{p}}} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} = \frac{1}{1.00\ \Omega} + \frac{1}{6.00\ \Omega} + \frac{1}{13.0\ \Omega}.$$

Thus,

Equation:

$$\frac{1}{R_{\text{p}}} = \frac{1.00}{\Omega} + \frac{0.1667}{\Omega} + \frac{0.07692}{\Omega} = \frac{1.2436}{\Omega}.$$

(Note that in these calculations, each intermediate answer is shown with an extra digit.)

We must invert this to find the total resistance R_{p}. This yields

Equation:

$$R_{\text{p}} = \frac{1}{1.2436}\Omega = 0.8041\ \Omega.$$

The total resistance with the correct number of significant digits is $R_{\text{p}} = 0.804\ \Omega$.

Discussion for (a)

R_{p} is, as predicted, less than the smallest individual resistance.

Strategy and Solution for (b)

The total current can be found from Ohm’s law, substituting R_{p} for the total resistance. This gives

Equation:

$$I = \frac{V}{R_{\text{p}}} = \frac{12.0\ \text{V}}{0.8041\ \Omega} = 14.92\ \text{A}.$$

Discussion for (b)
Current I for each device is much larger than for the same devices connected in series (see the previous example). A circuit with parallel connections has a smaller total resistance than the resistors connected in series.
Strategy and Solution for (c)
The individual currents are easily calculated from Ohm's law, since each resistor gets the full voltage. Thus,
Equation:

$$I_1 = \frac{V}{R_1} = \frac{12.0\ \text{V}}{1.00\,\Omega} = 12.0\ \text{A}.$$

Similarly,
Equation:

$$I_2 = \frac{V}{R_2} = \frac{12.0\ \text{V}}{6.00\,\Omega} = 2.00\ \text{A}$$

and
Equation:

$$I_3 = \frac{V}{R_3} = \frac{12.0\ \text{V}}{13.0\,\Omega} = 0.92\ \text{A}.$$

Discussion for (c)
The total current is the sum of the individual currents:
Equation:

$$I_1 + I_2 + I_3 = 14.92\ \text{A}.$$

This is consistent with conservation of charge.
Strategy and Solution for (d)
The power dissipated by each resistor can be found using any of the equations relating power to current, voltage, and resistance, since all three

are known. Let us use $P = \frac{V^2}{R}$, since each resistor gets full voltage. Thus,

Equation:

$$P_1 = \frac{V^2}{R_1} = \frac{(12.0\ \text{V})^2}{1.00\ \Omega} = 144\ \text{W}.$$

Similarly,

Equation:

$$P_2 = \frac{V^2}{R_2} = \frac{(12.0\ \text{V})^2}{6.00\ \Omega} = 24.0\ \text{W}$$

and

Equation:

$$P_3 = \frac{V^2}{R_3} = \frac{(12.0\ \text{V})^2}{13.0\ \Omega} = 11.1\ \text{W}.$$

Discussion for (d)

The power dissipated by each resistor is considerably higher in parallel than when connected in series to the same voltage source.

Strategy and Solution for (e)

The total power can also be calculated in several ways. Choosing $P = \text{IV}$, and entering the total current, yields

Equation:

$$P = \text{IV} = (14.92\ \text{A})(12.0\ \text{V}) = 179\ \text{W}.$$

Discussion for (e)

Total power dissipated by the resistors is also 179 W:

Equation:

$$P_1 + P_2 + P_3 = 144\ \text{W} + 24.0\ \text{W} + 11.1\ \text{W} = 179\ \text{W}.$$

This is consistent with the law of conservation of energy.

Overall Discussion

Note that both the currents and powers in parallel connections are greater than for the same devices in series.

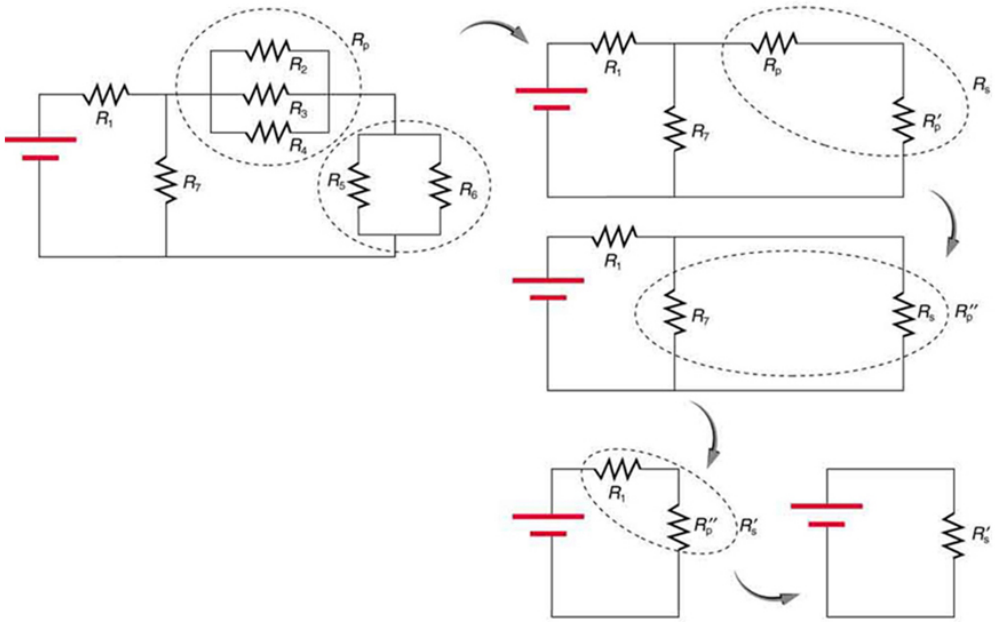
Note:
Major Features of Resistors in Parallel

1. Parallel resistance is found from $\frac{1}{R_{\text{p}}} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \ldots$, and it is smaller than any individual resistance in the combination.
2. Each resistor in parallel has the same full voltage of the source applied to it. (Power distribution systems most often use parallel connections to supply the myriad devices served with the same voltage and to allow them to operate independently.)
3. Parallel resistors do not each get the total current; they divide it.

Combinations of Series and Parallel

More complex connections of resistors are sometimes just combinations of series and parallel. These are commonly encountered, especially when wire resistance is considered. In that case, wire resistance is in series with other resistances that are in parallel.

Combinations of series and parallel can be reduced to a single equivalent resistance using the technique illustrated in [link]. Various parts are identified as either series or parallel, reduced to their equivalents, and further reduced until a single resistance is left. The process is more time consuming than difficult.

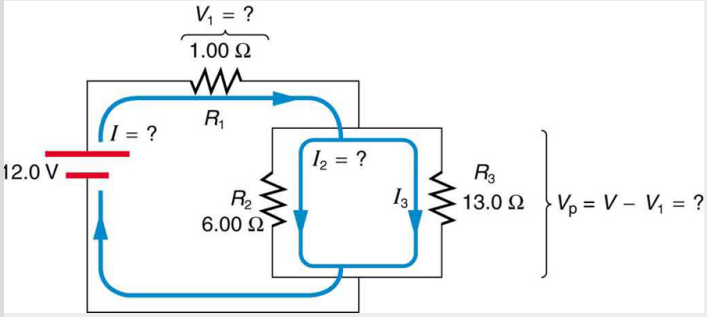


This combination of seven resistors has both series and parallel parts. Each is identified and reduced to an equivalent resistance, and these are further reduced until a single equivalent resistance is reached.

The simplest combination of series and parallel resistance, shown in [link], is also the most instructive, since it is found in many applications. For example, R_1 could be the resistance of wires from a car battery to its electrical devices, which are in parallel. R_2 and R_3 could be the starter motor and a passenger compartment light. We have previously assumed that wire resistance is negligible, but, when it is not, it has important effects, as the next example indicates.

Example:
Calculating Resistance, IR Drop, Current, and Power Dissipation: Combining Series and Parallel Circuits
[link] shows the resistors from the previous two examples wired in a different way—a combination of series and parallel. We can consider R_1 to be the resistance of wires leading to R_2 and R_3. (a) Find the total

resistance. (b) What is the IR drop in R_1? (c) Find the current I_2 through R_2. (d) What power is dissipated by R_2?



These three resistors are connected to a voltage source so that R_2 and R_3 are in parallel with one another and that combination is in series with R_1.

Strategy and Solution for (a)

To find the total resistance, we note that R_2 and R_3 are in parallel and their combination R_p is in series with R_1. Thus the total (equivalent) resistance of this combination is

Equation:

$$R_\text{tot} = R_1 + R_\text{p}.$$

First, we find R_p using the equation for resistors in parallel and entering known values:

Equation:

$$\frac{1}{R_\text{p}} = \frac{1}{R_2} + \frac{1}{R_3} = \frac{1}{6.00\ \Omega} + \frac{1}{13.0\ \Omega} = \frac{0.2436}{\Omega}.$$

Inverting gives

Equation:

$$R_\text{p} = \frac{1}{0.2436}\Omega = 4.11\ \Omega.$$

So the total resistance is

Equation:

$$R_{\text{tot}} = R_1 + R_{\text{p}} = 1.00\ \Omega + 4.11\ \Omega = 5.11\ \Omega.$$

Discussion for (a)
The total resistance of this combination is intermediate between the pure series and pure parallel values (20.0 Ω and 0.804 Ω, respectively) found for the same resistors in the two previous examples.
Strategy and Solution for (b)
To find the IR drop in R_1, we note that the full current I flows through R_1. Thus its IR drop is
Equation:

$$V_1 = \text{IR}_1.$$

We must find I before we can calculate V_1. The total current I is found using Ohm's law for the circuit. That is,
Equation:

$$I = \frac{V}{R_{\text{tot}}} = \frac{12.0\ \text{V}}{5.11\ \Omega} = 2.35\ \text{A}.$$

Entering this into the expression above, we get
Equation:

$$V_1 = \text{IR}_1 = (2.35\ \text{A})(1.00\ \Omega) = 2.35\ \text{V}.$$

Discussion for (b)
The voltage applied to R_2 and R_3 is less than the total voltage by an amount V_1. When wire resistance is large, it can significantly affect the operation of the devices represented by R_2 and R_3.
Strategy and Solution for (c)
To find the current through R_2, we must first find the voltage applied to it. We call this voltage V_{p}, because it is applied to a parallel combination of resistors. The voltage applied to both R_2 and R_3 is reduced by the amount V_1, and so it is
Equation:

$$V_{\text{p}} = V - V_1 = 12.0\ \text{V} - 2.35\ \text{V} = 9.65\ \text{V}.$$

Now the current I_2 through resistance R_2 is found using Ohm's law:
Equation:

$$I_2 = \frac{V_{\text{p}}}{R_2} = \frac{9.65\ \text{V}}{6.00\ \Omega} = 1.61\ \text{A}.$$

Discussion for (c)
The current is less than the 2.00 A that flowed through R_2 when it was connected in parallel to the battery in the previous parallel circuit example.
Strategy and Solution for (d)
The power dissipated by R_2 is given by
Equation:

$$P_2 = (I_2)^2 R_2 = (1.61\ \text{A})^2 (6.00\ \Omega) = 15.5\ \text{W}.$$

Discussion for (d)
The power is less than the 24.0 W this resistor dissipated when connected in parallel to the 12.0-V source.

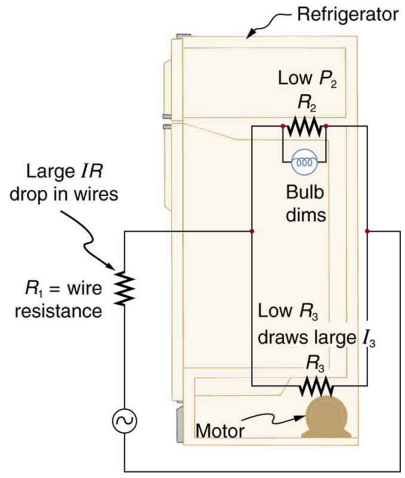
Practical Implications

One implication of this last example is that resistance in wires reduces the current and power delivered to a resistor. If wire resistance is relatively large, as in a worn (or a very long) extension cord, then this loss can be significant. If a large current is drawn, the IR drop in the wires can also be significant.

For example, when you are rummaging in the refrigerator and the motor comes on, the refrigerator light dims momentarily. Similarly, you can see the passenger compartment light dim when you start the engine of your car (although this may be due to resistance inside the battery itself).

What is happening in these high-current situations is illustrated in [link]. The device represented by R_3 has a very low resistance, and so when it is

switched on, a large current flows. This increased current causes a larger IR drop in the wires represented by R_1, reducing the voltage across the light bulb (which is R_2), which then dims noticeably.



Why do lights dim when a large appliance is switched on? The answer is that the large current the appliance motor draws causes a significant IR drop in the wires and reduces the voltage across the light.

Exercise:
Check Your Understanding

Problem:

Can any arbitrary combination of resistors be broken down into series and parallel combinations? See if you can draw a circuit diagram of resistors that cannot be broken down into combinations of series and parallel.

Solution:

No, there are many ways to connect resistors that are not combinations of series and parallel, including loops and junctions. In such cases Kirchhoff's rules, to be introduced in Kirchhoff's Rules, will allow you to analyze the circuit.

Note:
Problem-Solving Strategies for Series and Parallel Resistors

1. Draw a clear circuit diagram, labeling all resistors and voltage sources. This step includes a list of the knowns for the problem, since they are labeled in your circuit diagram.
2. Identify exactly what needs to be determined in the problem (identify the unknowns). A written list is useful.
3. Determine whether resistors are in series, parallel, or a combination of both series and parallel. Examine the circuit diagram to make this assessment. Resistors are in series if the same current must pass sequentially through them.
4. Use the appropriate list of major features for series or parallel connections to solve for the unknowns. There is one list for series and another for parallel. If your problem has a combination of series and parallel, reduce it in steps by considering individual groups of series or parallel connections, as done in this module and the examples. Special note: When finding R_{p}, the reciprocal must be taken with care.
5. Check to see whether the answers are reasonable and consistent. Units and numerical results must be reasonable. Total series resistance

should be greater, whereas total parallel resistance should be smaller, for example. Power should be greater for the same devices in parallel compared with series, and so on.

Section Summary

- The total resistance of an electrical circuit with resistors wired in a series is the sum of the individual resistances: $R_{\text{s}} = R_1 + R_2 + R_3 + ...$
- Each resistor in a series circuit has the same amount of current flowing through it.
- The voltage drop, or power dissipation, across each individual resistor in a series is different, and their combined total adds up to the power source input.
- The total resistance of an electrical circuit with resistors wired in parallel is less than the lowest resistance of any of the components and can be determined using the formula:
  Equation:

$$\frac{1}{R_{\text{p}}} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + ...$$

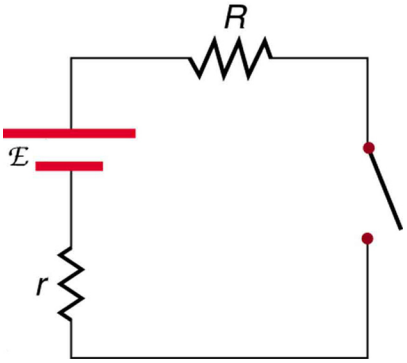
- Each resistor in a parallel circuit has the same full voltage of the source applied to it.
- The current flowing through each resistor in a parallel circuit is different, depending on the resistance.
- If a more complex connection of resistors is a combination of series and parallel, it can be reduced to a single equivalent resistance by identifying its various parts as series or parallel, reducing each to its equivalent, and continuing until a single resistance is eventually reached.

Conceptual Questions

Exercise:

Problem:

A switch has a variable resistance that is nearly zero when closed and extremely large when open, and it is placed in series with the device it controls. Explain the effect the switch in [link] has on current when open and when closed.



A switch is ordinarily in series with a resistance and voltage source. Ideally, the switch has nearly zero resistance when closed but has an extremely large resistance when open. (Note that in this diagram, the script E represents the voltage (or electromotive force) of the battery.)

Exercise:

Problem: What is the voltage across the open switch in [link]?

Exercise:

Problem:

There is a voltage across an open switch, such as in [link]. Why, then, is the power dissipated by the open switch small?

Exercise:

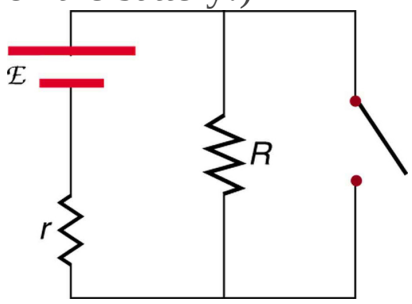
Problem:

Why is the power dissipated by a closed switch, such as in [link], small?

Exercise:

Problem:

A student in a physics lab mistakenly wired a light bulb, battery, and switch as shown in [link]. Explain why the bulb is on when the switch is open, and off when the switch is closed. (Do not try this—it is hard on the battery!)



A wiring mistake put this switch in parallel with the device represented by R. (Note that in this diagram, the script E represents the voltage (or

electromotive force) of the battery.)

Exercise:

Problem:

Knowing that the severity of a shock depends on the magnitude of the current through your body, would you prefer to be in series or parallel with a resistance, such as the heating element of a toaster, if shocked by it? Explain.

Exercise:

Problem:

Would your headlights dim when you start your car’s engine if the wires in your automobile were superconductors? (Do not neglect the battery’s internal resistance.) Explain.

Exercise:

Problem:

Some strings of holiday lights are wired in series to save wiring costs. An old version utilized bulbs that break the electrical connection, like an open switch, when they burn out. If one such bulb burns out, what happens to the others? If such a string operates on 120 V and has 40 identical bulbs, what is the normal operating voltage of each? Newer versions use bulbs that short circuit, like a closed switch, when they burn out. If one such bulb burns out, what happens to the others? If such a string operates on 120 V and has 39 remaining identical bulbs, what is then the operating voltage of each?

Exercise:

Problem:

If two household lightbulbs rated 60 W and 100 W are connected in series to household power, which will be brighter? Explain.

Exercise:

Problem:

Suppose you are doing a physics lab that asks you to put a resistor into a circuit, but all the resistors supplied have a larger resistance than the requested value. How would you connect the available resistances to attempt to get the smaller value asked for?

Exercise:

Problem:

Before World War II, some radios got power through a “resistance cord” that had a significant resistance. Such a resistance cord reduces the voltage to a desired level for the radio’s tubes and the like, and it saves the expense of a transformer. Explain why resistance cords become warm and waste energy when the radio is on.

Exercise:

Problem:

Some light bulbs have three power settings (not including zero), obtained from multiple filaments that are individually switched and wired in parallel. What is the minimum number of filaments needed for three power settings?

Problem Exercises

Note: Data taken from figures can be assumed to be accurate to three significant digits.

Exercise:

Problem:

(a) What is the resistance of ten $275\text{-}\Omega$ resistors connected in series? (b) In parallel?

Solution:

(a) $2.75\ \text{k}\Omega$

(b) $27.5\ \Omega$

Exercise:

Problem:

(a) What is the resistance of a 1.00×10^2 -Ω, a $2.50\text{-k}\Omega$, and a $4.00\text{-k}\Omega$ resistor connected in series? (b) In parallel?

Exercise:

Problem:

What are the largest and smallest resistances you can obtain by connecting a $36.0\text{-}\Omega$, a $50.0\text{-}\Omega$, and a $700\text{-}\Omega$ resistor together?

Solution:

(a) $786\ \Omega$

(b) $20.3\ \Omega$

Exercise:

Problem:

An 1800-W toaster, a 1400-W electric frying pan, and a 75-W lamp are plugged into the same outlet in a 15-A, 120-V circuit. (The three devices are in parallel when plugged into the same socket.). (a) What current is drawn by each device? (b) Will this combination blow the 15-A fuse?

Exercise:

Problem:

Your car’s 30.0-W headlight and 2.40-kW starter are ordinarily connected in parallel in a 12.0-V system. What power would one headlight and the starter consume if connected in series to a 12.0-V battery? (Neglect any other resistance in the circuit and any change in resistance in the two devices.)

Solution:

29.6 W

Exercise:

Problem:

(a) Given a 48.0-V battery and $24.0\text{-}\Omega$ and $96.0\text{-}\Omega$ resistors, find the current and power for each when connected in series. (b) Repeat when the resistances are in parallel.

Exercise:

Problem:

Referring to the example combining series and parallel circuits and [link], calculate I_3 in the following two different ways: (a) from the known values of I and I_2; (b) using Ohm’s law for R_3. In both parts explicitly show how you follow the steps in the Problem-Solving Strategies for Series and Parallel Resistors.

Solution:

(a) 0.74 A

(b) 0.742 A

Exercise:

Problem:

Referring to [link]: (a) Calculate P_3 and note how it compares with P_3 found in the first two example problems in this module. (b) Find the total power supplied by the source and compare it with the sum of the powers dissipated by the resistors.

Exercise:

Problem:

Refer to [link] and the discussion of lights dimming when a heavy appliance comes on. (a) Given the voltage source is 120 V, the wire resistance is $0.400\ \Omega$, and the bulb is nominally 75.0 W, what power will the bulb dissipate if a total of 15.0 A passes through the wires when the motor comes on? Assume negligible change in bulb resistance. (b) What power is consumed by the motor?

Solution:

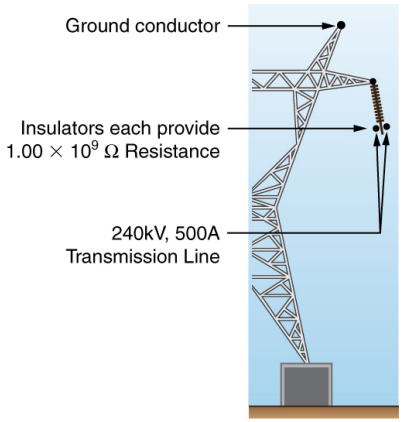
(a) 60.8 W

(b) 3.18 kW

Exercise:

Problem:

A 240-kV power transmission line carrying 5.00×10^2 A is hung from grounded metal towers by ceramic insulators, each having a 1.00×10^9 -Ω resistance. [link]. (a) What is the resistance to ground of 100 of these insulators? (b) Calculate the power dissipated by 100 of them. (c) What fraction of the power carried by the line is this? Explicitly show how you follow the steps in the Problem-Solving Strategies for Series and Parallel Resistors.



High-voltage (240-kV) transmission line carrying 5.00×10^2 A is hung from a grounded metal transmission tower. The row of ceramic insulators provide $1.00 \times 10^9 \Omega$ of resistance each.

Exercise:

Problem:

Show that if two resistors R_1 and R_2 are combined and one is much greater than the other ($R_1 >> R_2$): (a) Their series resistance is very nearly equal to the greater resistance R_1. (b) Their parallel resistance is very nearly equal to smaller resistance R_2.

Solution:

(a)
$$R_s = R_1 + R_2$$
$$\Rightarrow R_s \approx R_1 (R_1 >> R_2)$$

(b) $\frac{1}{R_{\text{p}}} = \frac{1}{R_1} + \frac{1}{R_2} = \frac{R_1+R_2}{R_1R_2}$,

so that

$R_{\text{p}} = \frac{R_1R_2}{R_1+R_2} \approx \frac{R_1R_2}{R_1} = R_2 (R_1 >> R_2)$.

Exercise:

Problem: Unreasonable Results

Two resistors, one having a resistance of $145\ \Omega$, are connected in parallel to produce a total resistance of $150\ \Omega$. (a) What is the value of the second resistance? (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Exercise:

Problem: Unreasonable Results

Two resistors, one having a resistance of $900\ \text{k}\Omega$, are connected in series to produce a total resistance of $0.500\ \text{M}\Omega$. (a) What is the value of the second resistance? (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Solution:

(a) $-400\ \text{k}\Omega$

(b) Resistance cannot be negative.

(c) Series resistance is said to be less than one of the resistors, but it must be greater than any of the resistors.

Glossary

series

a sequence of resistors or other components wired into a circuit one after the other

resistor
: a component that provides resistance to the current flowing through an electrical circuit

resistance
: causing a loss of electrical power in a circuit

Ohm's law
: the relationship between current, voltage, and resistance within an electrical circuit: $V = \mathrm{IR}$

voltage
: the electrical potential energy per unit charge; electric pressure created by a power source, such as a battery

voltage drop
: the loss of electrical power as a current travels through a resistor, wire or other component

current
: the flow of charge through an electric circuit past a given point of measurement

Joule's law
: the relationship between potential electrical power, voltage, and resistance in an electrical circuit, given by: $P_e = \mathrm{IV}$

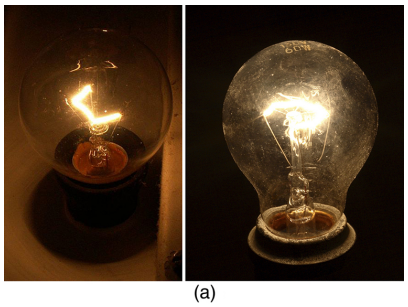
parallel
: the wiring of resistors or other components in an electrical circuit such that each component receives an equal voltage from the power source; often pictured in a ladder-shaped diagram, with each component on a rung of the ladder

Electric Power and Energy

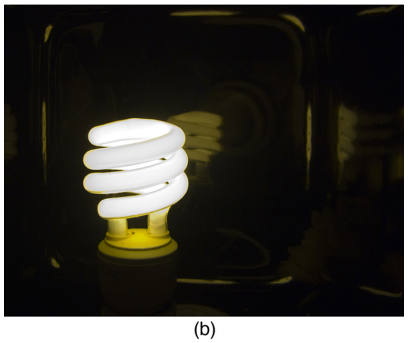
- Calculate the power dissipated by a resistor and power supplied by a power supply.
- Calculate the cost of electricity under various circumstances.

Power in Electric Circuits

Power is associated by many people with electricity. Knowing that power is the rate of energy use or energy conversion, what is the expression for **electric power**? Power transmission lines might come to mind. We also think of lightbulbs in terms of their power ratings in watts. Let us compare a 25-W bulb with a 60-W bulb. (See [link](a).) Since both operate on the same voltage, the 60-W bulb must draw more current to have a greater power rating. Thus the 60-W bulb's resistance must be lower than that of a 25-W bulb. If we increase voltage, we also increase power. For example, when a 25-W bulb that is designed to operate on 120 V is connected to 240 V, it briefly glows very brightly and then burns out. Precisely how are voltage, current, and resistance related to electric power?



(a)



(b)

(a) Which of these lightbulbs, the 25-W bulb (upper left) or the 60-W bulb (upper right), has the higher resistance? Which draws more current? Which uses the most energy? Can you tell from the color that the 25-W filament is cooler? Is the brighter bulb a different color and if so why? (credits: Dickbauch, Wikimedia Commons; Greg Westfall, Flickr) (b) This compact fluorescent light (CFL) puts out the same intensity of light as the 60-W bulb, but at 1/4 to 1/10 the input power. (credit: dbgg1979, Flickr)

Electric energy depends on both the voltage involved and the charge moved. This is expressed most simply as $\mathrm{PE} = \mathrm{qV}$, where q is the charge moved and V is the voltage (or more precisely, the potential difference the

charge moves through). Power is the rate at which energy is moved, and so electric power is

Equation:

$$P = \frac{\text{PE}}{t} = \frac{\text{qV}}{t}.$$

Recognizing that current is $I = q/t$ (note that $\Delta t = t$ here), the expression for power becomes

Equation:

$$P = \text{IV}.$$

Electric power (P) is simply the product of current times voltage. Power has familiar units of watts. Since the SI unit for potential energy (PE) is the joule, power has units of joules per second, or watts. Thus, $1\ \text{A} \cdot \text{V} = 1\ \text{W}$. For example, cars often have one or more auxiliary power outlets with which you can charge a cell phone or other electronic devices. These outlets may be rated at 20 A, so that the circuit can deliver a maximum power $P = \text{IV} = (20\ \text{A})(12\ \text{V}) = 240\ \text{W}$. In some applications, electric power may be expressed as volt-amperes or even kilovolt-amperes ($1\ \text{kA} \cdot \text{V} = 1\ \text{kW}$).

To see the relationship of power to resistance, we combine Ohm’s law with $P = \text{IV}$. Substituting $I = \text{V}/\text{R}$ gives $P = (V/R)V = V^2/R$. Similarly, substituting $V = \text{IR}$ gives $P = I(\text{IR}) = I^2R$. Three expressions for electric power are listed together here for convenience:

Equation:

$$P = \text{IV}$$

Equation:

$$P = \frac{V^2}{R}$$

Equation:

$$P = I^2R.$$

Note that the first equation is always valid, whereas the other two can be used only for resistors. In a simple circuit, with one voltage source and a single resistor, the power supplied by the voltage source and that dissipated by the resistor are identical. (In more complicated circuits, P can be the power dissipated by a single device and not the total power in the circuit.)

Different insights can be gained from the three different expressions for electric power. For example, $P = V^2/R$ implies that the lower the resistance connected to a given voltage source, the greater the power delivered. Furthermore, since voltage is squared in $P = V^2/R$, the effect of applying a higher voltage is perhaps greater than expected. Thus, when the voltage is doubled to a 25-W bulb, its power nearly quadruples to about 100 W, burning it out. If the bulb's resistance remained constant, its power would be exactly 100 W, but at the higher temperature its resistance is higher, too.

Example:
Calculating Power Dissipation and Current: Hot and Cold Power
(a) Consider the examples given in Ohm's Law: Resistance and Simple Circuits and Resistance and Resistivity. Then find the power dissipated by the car headlight in these examples, both when it is hot and when it is cold.
(b) What current does it draw when cold?
Strategy for (a)
For the hot headlight, we know voltage and current, so we can use $P = \text{IV}$ to find the power. For the cold headlight, we know the voltage and resistance, so we can use $P = V^2/R$ to find the power.
Solution for (a)
Entering the known values of current and voltage for the hot headlight, we obtain
Equation:

$$P = \text{IV} = (2.50\ \text{A})(12.0\ \text{V}) = 30.0\ \text{W}.$$

The cold resistance was $0.350\ \Omega$, and so the power it uses when first switched on is

Equation:

$$P = \frac{V^2}{R} = \frac{(12.0\ \text{V})^2}{0.350\ \Omega} = 411\ \text{W}.$$

Discussion for (a)

The 30 W dissipated by the hot headlight is typical. But the 411 W when cold is surprisingly higher. The initial power quickly decreases as the bulb's temperature increases and its resistance increases.

Strategy and Solution for (b)

The current when the bulb is cold can be found several different ways. We rearrange one of the power equations, $P = I^2R$, and enter known values, obtaining

Equation:

$$I = \sqrt{\frac{P}{R}} = \sqrt{\frac{411\ \text{W}}{0.350\ \Omega}} = 34.3\ \text{A}.$$

Discussion for (b)

The cold current is remarkably higher than the steady-state value of 2.50 A, but the current will quickly decline to that value as the bulb's temperature increases. Most fuses and circuit breakers (used to limit the current in a circuit) are designed to tolerate very high currents briefly as a device comes on. In some cases, such as with electric motors, the current remains high for several seconds, necessitating special "slow blow" fuses.

The Cost of Electricity

The more electric appliances you use and the longer they are left on, the higher your electric bill. This familiar fact is based on the relationship between energy and power. You pay for the energy used. Since $P = E/t$, we see that

Equation:

$$E = \mathrm{Pt}$$

is the energy used by a device using power P for a time interval t. For example, the more lightbulbs burning, the greater P used; the longer they are on, the greater t is. The energy unit on electric bills is the kilowatt-hour ($\mathrm{kW} \cdot \mathrm{h}$), consistent with the relationship $E = \mathrm{Pt}$. It is easy to estimate the cost of operating electric appliances if you have some idea of their power consumption rate in watts or kilowatts, the time they are on in hours, and the cost per kilowatt-hour for your electric utility. Kilowatt-hours, like all other specialized energy units such as food calories, can be converted to joules. You can prove to yourself that $1\ \mathrm{kW} \cdot \mathrm{h} = 3.6 \times 10^6\ \mathrm{J}$.

The electrical energy (E) used can be reduced either by reducing the time of use or by reducing the power consumption of that appliance or fixture. This will not only reduce the cost, but it will also result in a reduced impact on the environment. Improvements to lighting are some of the fastest ways to reduce the electrical energy used in a home or business. About 20% of a home's use of energy goes to lighting, while the number for commercial establishments is closer to 40%. Fluorescent lights are about four times more efficient than incandescent lights—this is true for both the long tubes and the compact fluorescent lights (CFL). (See [link](b).) Thus, a 60-W incandescent bulb can be replaced by a 15-W CFL, which has the same brightness and color. CFLs have a bent tube inside a globe or a spiral-shaped tube, all connected to a standard screw-in base that fits standard incandescent light sockets. (Original problems with color, flicker, shape, and high initial investment for CFLs have been addressed in recent years.) The heat transfer from these CFLs is less, and they last up to 10 times longer. The significance of an investment in such bulbs is addressed in the next example. New white LED lights (which are clusters of small LED bulbs) are even more efficient (twice that of CFLs) and last 5 times longer than CFLs. However, their cost is still high.

Note:
Making Connections: Energy, Power, and Time

The relationship $E = \mathrm{Pt}$ is one that you will find useful in many different contexts. The energy your body uses in exercise is related to the power level and duration of your activity, for example. The amount of heating by a power source is related to the power level and time it is applied. Even the radiation dose of an X-ray image is related to the power and time of exposure.

Example:
Calculating the Cost Effectiveness of Compact Fluorescent Lights (CFL)
If the cost of electricity in your area is 12 cents per kWh, what is the total cost (capital plus operation) of using a 60-W incandescent bulb for 1000 hours (the lifetime of that bulb) if the bulb cost 25 cents? (b) If we replace this bulb with a compact fluorescent light that provides the same light output, but at one-quarter the wattage, and which costs $1.50 but lasts 10 times longer (10,000 hours), what will that total cost be?
Strategy
To find the operating cost, we first find the energy used in kilowatt-hours and then multiply by the cost per kilowatt-hour.
Solution for (a)
The energy used in kilowatt-hours is found by entering the power and time into the expression for energy:
Equation:

$$E = \mathrm{Pt} = (60\ \mathrm{W})(1000\ \mathrm{h}) = 60{,}000\ \mathrm{W \cdot h}.$$

In kilowatt-hours, this is
Equation:

$$E = 60.0\ \mathrm{kW \cdot h}.$$

Now the electricity cost is
Equation:

$$\mathrm{cost} = (60.0\ \mathrm{kW \cdot h})(\$0.12/\mathrm{kW \cdot h}) = \$7.20.$$

The total cost will be $7.20 for 1000 hours (about one-half year at 5 hours per day).

Solution for (b)

Since the CFL uses only 15 W and not 60 W, the electricity cost will be $7.20/4 = $1.80. The CFL will last 10 times longer than the incandescent, so that the investment cost will be 1/10 of the bulb cost for that time period of use, or 0.1($1.50) = $0.15. Therefore, the total cost will be $1.95 for 1000 hours.

Discussion

Therefore, it is much cheaper to use the CFLs, even though the initial investment is higher. The increased cost of labor that a business must include for replacing the incandescent bulbs more often has not been figured in here.

Note:

Making Connections: Take-Home Experiment—Electrical Energy Use Inventory

1) Make a list of the power ratings on a range of appliances in your home or room. Explain why something like a toaster has a higher rating than a digital clock. Estimate the energy consumed by these appliances in an average day (by estimating their time of use). Some appliances might only state the operating current. If the household voltage is 120 V, then use $P = \text{IV}$. 2) Check out the total wattage used in the rest rooms of your school's floor or building. (You might need to assume the long fluorescent lights in use are rated at 32 W.) Suppose that the building was closed all weekend and that these lights were left on from 6 p.m. Friday until 8 a.m. Monday. What would this oversight cost? How about for an entire year of weekends?

Section Summary

- Electric power P is the rate (in watts) that energy is supplied by a source or dissipated by a device.

- Three expressions for electrical power are
  Equation:

$$P = \mathrm{IV},$$

  Equation:

$$P = \frac{V^2}{R},$$

  and
  Equation:

$$P = I^2R.$$

- The energy used by a device with a power P over a time t is $E = \mathrm{Pt}$.

Conceptual Questions

Exercise:

Problem:

Why do incandescent lightbulbs grow dim late in their lives, particularly just before their filaments break?

Exercise:

Problem:

The power dissipated in a resistor is given by $P = V^2/R$, which means power decreases if resistance increases. Yet this power is also given by $P = I^2R$, which means power increases if resistance increases. Explain why there is no contradiction here.

Problem Exercises

Exercise:

Problem:

What is the power of a 1.00×10^2 MV lightning bolt having a current of 2.00×10^4 A?

Solution:

2.00×10^{12} W

Exercise:

Problem:

What power is supplied to the starter motor of a large truck that draws 250 A of current from a 24.0-V battery hookup?

Exercise:

Problem:

A charge of 4.00 C of charge passes through a pocket calculator's solar cells in 4.00 h. What is the power output, given the calculator's voltage output is 3.00 V? (See [link].)



The strip of solar cells just above the keys of this calculator convert

light to electricity to supply its energy needs. (credit: Evan-Amos, Wikimedia Commons)

Exercise:

Problem:

How many watts does a flashlight that has 6.00×10^2 C pass through it in 0.500 h use if its voltage is 3.00 V?

Exercise:

Problem:

Find the power dissipated in each of these extension cords: (a) an extension cord having a 0.0600 - Ω resistance and through which 5.00 A is flowing; (b) a cheaper cord utilizing thinner wire and with a resistance of $0.300\ \Omega$.

Solution:

(a) 1.50 W

(b) 7.50 W

Exercise:

Problem:

Verify that the units of a volt-ampere are watts, as implied by the equation $P = \mathrm{IV}$.

Exercise:

Problem:

Show that the units $1\ \mathrm{V}^2/\Omega = 1\mathrm{W}$, as implied by the equation $P = V^2/R$.

Solution:

$$\frac{V^2}{\Omega} = \frac{V^2}{\mathrm{V/A}} = \mathrm{AV} = \left(\frac{C}{s}\right)\left(\frac{J}{C}\right) = \frac{J}{s} = 1\ \mathrm{W}$$

Exercise:

Problem:

Show that the units $1\ \mathrm{A}^2 \cdot \Omega = 1\ \mathrm{W}$, as implied by the equation $P = I^2R$.

Exercise:

Problem:

Verify the energy unit equivalence that $1\ \mathrm{kW \cdot h} = 3.60{\times}10^6\ \mathrm{J}$.

Solution:

$$1\ \mathrm{kW \cdot h} = \left(\frac{1{\times}10^3\ \mathrm{J}}{1\ \mathrm{s}}\right)(1\ \mathrm{h})\left(\frac{3600\ \mathrm{s}}{1\ \mathrm{h}}\right) = 3.60 \times 10^6\ \mathrm{J}$$

Exercise:

Problem:

Electrons in an X-ray tube are accelerated through $1.00 \times 10^2\ \mathrm{kV}$ and directed toward a target to produce X-rays. Calculate the power of the electron beam in this tube if it has a current of 15.0 mA.

Exercise:

Problem:

An electric water heater consumes 5.00 kW for 2.00 h per day. What is the cost of running it for one year if electricity costs $12.0\ \text{cents}/\text{kW} \cdot \text{h}$? See [link].



On-demand electric hot water heater. Heat is supplied to water only when needed. (credit: aviddavid, Flickr)

Solution:

$438/y

Exercise:

Problem:

With a 1200-W toaster, how much electrical energy is needed to make a slice of toast (cooking time = 1 minute)? At $9.0\ \text{cents}/\text{kW} \cdot \text{h}$, how much does this cost?

Exercise:

Problem:

What would be the maximum cost of a CFL such that the total cost (investment plus operating) would be the same for both CFL and incandescent 60-W bulbs? Assume the cost of the incandescent bulb is 25 cents and that electricity costs 10 cents/kWh. Calculate the cost for 1000 hours, as in the cost effectiveness of CFL example.

Solution:

$6.25

Exercise:

Problem:

Some makes of older cars have 6.00-V electrical systems. (a) What is the hot resistance of a 30.0-W headlight in such a car? (b) What current flows through it?

Exercise:

Problem:

Alkaline batteries have the advantage of putting out constant voltage until very nearly the end of their life. How long will an alkaline battery rated at 1.00 $\text{A} \cdot \text{h}$ and 1.58 V keep a 1.00-W flashlight bulb burning?

Solution:

1.58 h

Exercise:

Problem:

A cauterizer, used to stop bleeding in surgery, puts out 2.00 mA at 15.0 kV. (a) What is its power output? (b) What is the resistance of the path?

Exercise:

Problem:

The average television is said to be on 6 hours per day. Estimate the yearly cost of electricity to operate 100 million TVs, assuming their power consumption averages 150 W and the cost of electricity averages 12.0 $\text{cents}/\text{kW} \cdot \text{h}$.

Solution:

$3.94 billion/year

Exercise:

Problem:

An old lightbulb draws only 50.0 W, rather than its original 60.0 W, due to evaporative thinning of its filament. By what factor is its diameter reduced, assuming uniform thinning along its length? Neglect any effects caused by temperature differences.

Exercise:

Problem:

00-gauge copper wire has a diameter of 9.266 mm. Calculate the power loss in a kilometer of such wire when it carries 1.00×10^2 A.

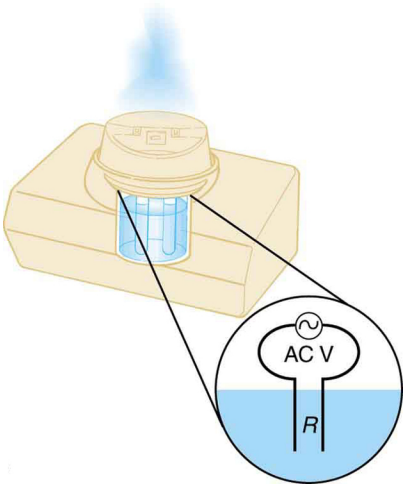
Solution:

25.5 W

Exercise:

Problem: Integrated Concepts

Cold vaporizers pass a current through water, evaporating it with only a small increase in temperature. One such home device is rated at 3.50 A and utilizes 120 V AC with 95.0% efficiency. (a) What is the vaporization rate in grams per minute? (b) How much water must you put into the vaporizer for 8.00 h of overnight operation? (See [link].)



This cold vaporizer passes current directly through water, vaporizing it directly with relatively little temperature increase.

Exercise:

Problem: Integrated Concepts

(a) What energy is dissipated by a lightning bolt having a 20,000-A current, a voltage of 1.00×10^2 MV, and a length of 1.00 ms? (b) What mass of tree sap could be raised from 18.0°C to its boiling point and then evaporated by this energy, assuming sap has the same thermal characteristics as water?

Solution:

(a) 2.00×10^9 J

(b) 769 kg

Exercise:

Problem: **Integrated Concepts**

What current must be produced by a 12.0-V battery-operated bottle warmer in order to heat 75.0 g of glass, 250 g of baby formula, and 3.00×10^2 g of aluminum from 20.0°C to 90.0°C in 5.00 min?

Exercise:

Problem: **Integrated Concepts**

How much time is needed for a surgical cauterizer to raise the temperature of 1.00 g of tissue from 37.0°C to 100°C and then boil away 0.500 g of water, if it puts out 2.00 mA at 15.0 kV? Ignore heat transfer to the surroundings.

Solution:

45.0 s

Exercise:

Problem: **Integrated Concepts**

Hydroelectric generators (see [link]) at Hoover Dam produce a maximum current of 8.00×10^3 A at 250 kV. (a) What is the power output? (b) The water that powers the generators enters and leaves the system at low speed (thus its kinetic energy does not change) but loses 160 m in altitude. How many cubic meters per second are needed, assuming 85.0% efficiency?



Hydroelectric generators
at the Hoover dam.
(credit: Jon Sullivan)

Exercise:

Problem: Integrated Concepts

(a) Assuming 95.0% efficiency for the conversion of electrical power by the motor, what current must the 12.0-V batteries of a 750-kg electric car be able to supply: (a) To accelerate from rest to 25.0 m/s in 1.00 min? (b) To climb a 2.00×10^2-m-high hill in 2.00 min at a constant 25.0-m/s speed while exerting 5.00×10^2 N of force to overcome air resistance and friction? (c) To travel at a constant 25.0-m/s speed, exerting a 5.00×10^2 N force to overcome air resistance and friction? See [link].



This REVAi, an electric

car, gets recharged on a street in London. (credit: Frank Hebbert)

Solution:

(a) 343 A

(b) 2.17×10^{3} A

(c) 1.10×10^{3} A

Exercise:

Problem: Integrated Concepts

A light-rail commuter train draws 630 A of 650-V DC electricity when accelerating. (a) What is its power consumption rate in kilowatts? (b) How long does it take to reach 20.0 m/s starting from rest if its loaded mass is 5.30×10^{4} kg, assuming 95.0% efficiency and constant power? (c) Find its average acceleration. (d) Discuss how the acceleration you found for the light-rail train compares to what might be typical for an automobile.

Exercise:

Problem: Integrated Concepts

(a) An aluminum power transmission line has a resistance of $0.0580\ \Omega/\text{km}$. What is its mass per kilometer? (b) What is the mass per kilometer of a copper line having the same resistance? A lower resistance would shorten the heating time. Discuss the practical limits to speeding the heating by lowering the resistance.

Solution:

(a) 1.23×10^3 kg

(b) 2.64×10^3 kg

Exercise:

Problem: **Integrated Concepts**

(a) An immersion heater utilizing 120 V can raise the temperature of a 1.00×10^2-g aluminum cup containing 350 g of water from 20.0°C to 95.0°C in 2.00 min. Find its resistance, assuming it is constant during the process. (b) A lower resistance would shorten the heating time. Discuss the practical limits to speeding the heating by lowering the resistance.

Exercise:

Problem: **Integrated Concepts**

(a) What is the cost of heating a hot tub containing 1500 kg of water from 10.0°C to 40.0°C, assuming 75.0% efficiency to account for heat transfer to the surroundings? The cost of electricity is $9\ \text{cents}/\text{kW} \cdot \text{h}$. (b) What current was used by the 220-V AC electric heater, if this took 4.00 h?

Exercise:

Problem: **Unreasonable Results**

(a) What current is needed to transmit 1.00×10^2 MW of power at 480 V? (b) What power is dissipated by the transmission lines if they have a 1.00 - Ω resistance? (c) What is unreasonable about this result? (d) Which assumptions are unreasonable, or which premises are inconsistent?

Solution:

(a) 2.08×10^5 A

(b) 4.33×10^4 MW

(c) The transmission lines dissipate more power than they are supposed to transmit.

(d) A voltage of 480 V is unreasonably low for a transmission voltage. Long-distance transmission lines are kept at much higher voltages (often hundreds of kilovolts) to reduce power losses.

Exercise:

Problem: Unreasonable Results

(a) What current is needed to transmit 1.00×10^2 MW of power at 10.0 kV? (b) Find the resistance of 1.00 km of wire that would cause a 0.0100% power loss. (c) What is the diameter of a 1.00-km-long copper wire having this resistance? (d) What is unreasonable about these results? (e) Which assumptions are unreasonable, or which premises are inconsistent?

Exercise:

Problem: Construct Your Own Problem

Consider an electric immersion heater used to heat a cup of water to make tea. Construct a problem in which you calculate the needed resistance of the heater so that it increases the temperature of the water and cup in a reasonable amount of time. Also calculate the cost of the electrical energy used in your process. Among the things to be considered are the voltage used, the masses and heat capacities involved, heat losses, and the time over which the heating takes place. Your instructor may wish for you to consider a thermal safety switch (perhaps bimetallic) that will halt the process before damaging temperatures are reached in the immersion unit.

Glossary

electric power

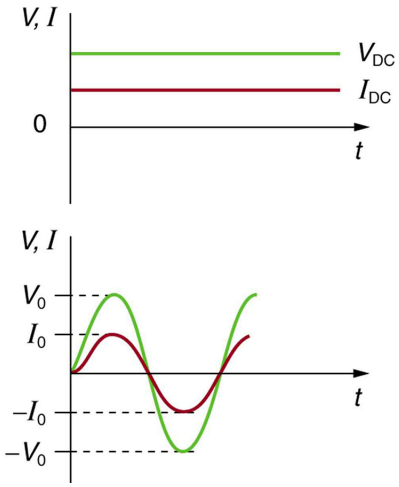
: the rate at which electrical energy is supplied by a source or dissipated by a device; it is the product of current times voltage

Alternating Current versus Direct Current

- Explain the differences and similarities between AC and DC current.
- Calculate rms voltage, current, and average power.
- Explain why AC current is used for power transmission.

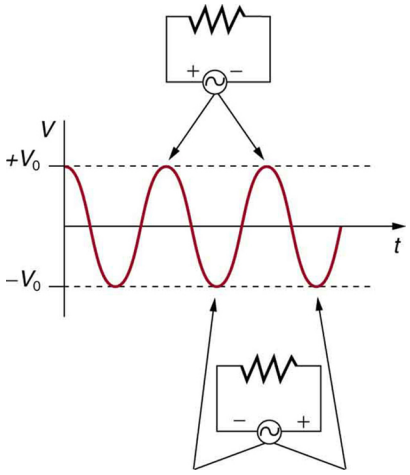
Alternating Current

Most of the examples dealt with so far, and particularly those utilizing batteries, have constant voltage sources. Once the current is established, it is thus also a constant. **Direct current** (DC) is the flow of electric charge in only one direction. It is the steady state of a constant-voltage circuit. Most well-known applications, however, use a time-varying voltage source. **Alternating current** (AC) is the flow of electric charge that periodically reverses direction. If the source varies periodically, particularly sinusoidally, the circuit is known as an alternating current circuit. Examples include the commercial and residential power that serves so many of our needs. [link] shows graphs of voltage and current versus time for typical DC and AC power. The AC voltages and frequencies commonly used in homes and businesses vary around the world.



(a) DC voltage and current are constant in time, once the

current is established. (b) A graph of voltage and current versus time for 60-Hz AC power. The voltage and current are sinusoidal and are in phase for a simple resistance circuit. The frequencies and peak voltages of AC sources differ greatly.



The potential difference V between the terminals of an AC voltage source fluctuates as

shown. The mathematical expression for V is given by $V = V_0 \sin 2\pi \mathrm{ft}$.

[link] shows a schematic of a simple circuit with an AC voltage source. The voltage between the terminals fluctuates as shown, with the **AC voltage** given by

Equation:

$$V = V_0 \sin 2\pi \mathrm{ft},$$

where V is the voltage at time t, V_0 is the peak voltage, and f is the frequency in hertz. For this simple resistance circuit, $I = \mathrm{V}/\mathrm{R}$, and so the **AC current** is

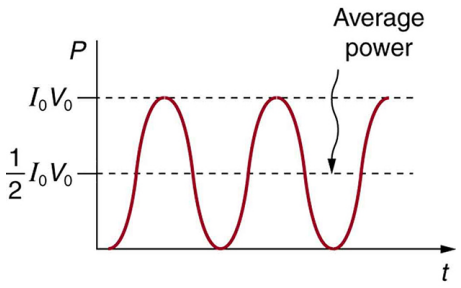
Equation:

$$I = I_0 \sin 2\pi \mathrm{ft},$$

where I is the current at time t, and $I_0 = V_0/\mathrm{R}$ is the peak current. For this example, the voltage and current are said to be in phase, as seen in [link](b).

Current in the resistor alternates back and forth just like the driving voltage, since $I = \mathrm{V}/\mathrm{R}$. If the resistor is a fluorescent light bulb, for example, it brightens and dims 120 times per second as the current repeatedly goes through zero. A 120-Hz flicker is too rapid for your eyes to detect, but if you wave your hand back and forth between your face and a fluorescent light, you will see a stroboscopic effect evidencing AC. The fact that the light output fluctuates means that the power is fluctuating. The power supplied is $P = \mathrm{IV}$. Using the expressions for I and V above, we see that the time dependence of power is $P = I_0 V_0 \sin^2 2\pi \mathrm{ft}$, as shown in [link].

Note:
Making Connections: Take-Home Experiment—AC/DC Lights
Wave your hand back and forth between your face and a fluorescent light bulb. Do you observe the same thing with the headlights on your car? Explain what you observe. *Warning: Do not look directly at very bright light.*



AC power as a function of time. Since the voltage and current are in phase here, their product is non-negative and fluctuates between zero and I_0V_0. Average power is $(1/2)I_0V_0$.

We are most often concerned with average power rather than its fluctuations—that 60-W light bulb in your desk lamp has an average power consumption of 60 W, for example. As illustrated in [link], the average power P_{ave} is
Equation:

$$P_{\text{ave}} = \frac{1}{2}I_0V_0.$$

This is evident from the graph, since the areas above and below the $(1/2)I_0V_0$ line are equal, but it can also be proven using trigonometric identities. Similarly, we define an average or **rms current** I_{rms} and average or **rms voltage** V_{rms} to be, respectively,
Equation:

$$I_{\text{rms}} = \frac{I_0}{\sqrt{2}}$$

and
Equation:

$$V_{\text{rms}} = \frac{V_0}{\sqrt{2}}.$$

where rms stands for root mean square, a particular kind of average. In general, to obtain a root mean square, the particular quantity is squared, its mean (or average) is found, and the square root is taken. This is useful for AC, since the average value is zero. Now,
Equation:

$$P_{\text{ave}} = I_{\text{rms}}V_{\text{rms}},$$

which gives
Equation:

$$P_{\text{ave}} = \frac{I_0}{\sqrt{2}} \cdot \frac{V_0}{\sqrt{2}} = \frac{1}{2}I_0V_0,$$

as stated above. It is standard practice to quote I_{rms}, V_{rms}, and P_{ave} rather than the peak values. For example, most household electricity is 120 V AC, which means that V_{rms} is 120 V. The common 10-A circuit breaker will interrupt a sustained I_{rms} greater than 10 A. Your 1.0-kW microwave oven

consumes $P_{\text{ave}} = 1.0$ kW, and so on. You can think of these rms and average values as the equivalent DC values for a simple resistive circuit.

To summarize, when dealing with AC, Ohm's law and the equations for power are completely analogous to those for DC, but rms and average values are used for AC. Thus, for AC, Ohm's law is written
Equation:

$$I_{\text{rms}} = \frac{V_{\text{rms}}}{R}.$$

The various expressions for AC power P_{ave} are
Equation:

$$P_{\text{ave}} = I_{\text{rms}} V_{\text{rms}},$$

Equation:

$$P_{\text{ave}} = \frac{V_{\text{rms}}^2}{R},$$

and
Equation:

$$P_{\text{ave}} = I_{\text{rms}}^2 R.$$

Example:
Peak Voltage and Power for AC
(a) What is the value of the peak voltage for 120-V AC power? (b) What is the peak power consumption rate of a 60.0-W AC light bulb?
Strategy
We are told that V_{rms} is 120 V and P_{ave} is 60.0 W. We can use $V_{\text{rms}} = \frac{V_0}{\sqrt{2}}$ to find the peak voltage, and we can manipulate the definition of power to

find the peak power from the given average power.

Solution for (a)

Solving the equation $V_{\text{rms}} = \frac{V_0}{\sqrt{2}}$ for the peak voltage V_0 and substituting the known value for V_{rms} gives

Equation:

$$V_0 = \sqrt{2}V_{\text{rms}} = 1.414(120\ \text{V}) = 170\ \text{V}.$$

Discussion for (a)

This means that the AC voltage swings from 170 V to –170 V and back 60 times every second. An equivalent DC voltage is a constant 120 V.

Solution for (b)

Peak power is peak current times peak voltage. Thus,

Equation:

$$P_0 = I_0V_0 = 2\left(\frac{1}{2}I_0V_0\right) = 2P_{\text{ave}}.$$

We know the average power is 60.0 W, and so

Equation:

$$P_0 = 2(60.0\ \text{W}) = 120\ \text{W}.$$

Discussion

So the power swings from zero to 120 W one hundred twenty times per second (twice each cycle), and the power averages 60 W.

Why Use AC for Power Distribution?

Most large power-distribution systems are AC. Moreover, the power is transmitted at much higher voltages than the 120-V AC (240 V in most parts of the world) we use in homes and on the job. Economies of scale make it cheaper to build a few very large electric power-generation plants than to build numerous small ones. This necessitates sending power long distances, and it is obviously important that energy losses en route be

minimized. High voltages can be transmitted with much smaller power losses than low voltages, as we shall see. (See [link].) For safety reasons, the voltage at the user is reduced to familiar values. The crucial factor is that it is much easier to increase and decrease AC voltages than DC, so AC is used in most large power distribution systems.



Power is distributed over large distances at high voltage to reduce power loss in the transmission lines. The voltages generated at the power plant are stepped up by passive devices called transformers (see Transformers) to 330,000 volts (or more in some places worldwide). At the point of use, the transformers reduce the voltage transmitted for safe residential and commercial use. (Credit: GeorgHH, Wikimedia Commons)

Example:
Power Losses Are Less for High-Voltage Transmission
(a) What current is needed to transmit 100 MW of power at 200 kV? (b) What is the power dissipated by the transmission lines if they have a resistance of 1.00 Ω? (c) What percentage of the power is lost in the transmission lines?
Strategy
We are given $P_{\text{ave}} = 100$ MW, $V_{\text{rms}} = 200$ kV, and the resistance of the lines is $R = 1.00\ \Omega$. Using these givens, we can find the current flowing (from $P = \text{IV}$) and then the power dissipated in the lines ($P = I^2R$), and we take the ratio to the total power transmitted.
Solution
To find the current, we rearrange the relationship $P_{\text{ave}} = I_{\text{rms}}V_{\text{rms}}$ and substitute known values. This gives
Equation:

$$I_{\text{rms}} = \frac{P_{\text{ave}}}{V_{\text{rms}}} = \frac{100 \times 10^6\ \text{W}}{200 \times 10^3\ \text{V}} = 500\ \text{A}.$$

Solution
Knowing the current and given the resistance of the lines, the power dissipated in them is found from $P_{\text{ave}} = I_{\text{rms}}^2R$. Substituting the known values gives
Equation:

$$P_{\text{ave}} = I_{\text{rms}}^2R = (500\ \text{A})^2(1.00\ \Omega) = 250\ \text{kW}.$$

Solution
The percent loss is the ratio of this lost power to the total or input power, multiplied by 100:
Equation:

$$\% \text{ loss} = \frac{250 \text{ kW}}{100 \text{ MW}} \times 100 = 0.250 \%.$$

Discussion

One-fourth of a percent is an acceptable loss. Note that if 100 MW of power had been transmitted at 25 kV, then a current of 4000 A would have been needed. This would result in a power loss in the lines of 16.0 MW, or 16.0% rather than 0.250%. The lower the voltage, the more current is needed, and the greater the power loss in the fixed-resistance transmission lines. Of course, lower-resistance lines can be built, but this requires larger and more expensive wires. If superconducting lines could be economically produced, there would be no loss in the transmission lines at all. But, as we shall see in a later chapter, there is a limit to current in superconductors, too. In short, high voltages are more economical for transmitting power, and AC voltage is much easier to raise and lower, so that AC is used in most large-scale power distribution systems.

It is widely recognized that high voltages pose greater hazards than low voltages. But, in fact, some high voltages, such as those associated with common static electricity, can be harmless. So it is not voltage alone that determines a hazard. It is not so widely recognized that AC shocks are often more harmful than similar DC shocks. Thomas Edison thought that AC shocks were more harmful and set up a DC power-distribution system in New York City in the late 1800s. There were bitter fights, in particular between Edison and George Westinghouse and Nikola Tesla, who were advocating the use of AC in early power-distribution systems. AC has prevailed largely due to transformers and lower power losses with high-voltage transmission.

Note:

PhET Explorations: Generator

Generate electricity with a bar magnet! Discover the physics behind the phenomena by exploring magnets and how you can use them to make a bulb light.

Generator

Section Summary

- Direct current (DC) is the flow of electric current in only one direction. It refers to systems where the source voltage is constant.
- The voltage source of an alternating current (AC) system puts out $V = V_0 \sin 2\pi \mathrm{ft}$, where V is the voltage at time t, V_0 is the peak voltage, and f is the frequency in hertz.
- In a simple circuit, $I = \mathrm{V}/\mathrm{R}$ and AC current is $I = I_0 \sin 2\pi \mathrm{ft}$, where I is the current at time t, and $I_0 = V_0/\mathrm{R}$ is the peak current.
- The average AC power is $P_{\text{ave}} = \frac{1}{2} I_0 V_0$.
- Average (rms) current I_{rms} and average (rms) voltage V_{rms} are $I_{\text{rms}} = \frac{I_0}{\sqrt{2}}$ and $V_{\text{rms}} = \frac{V_0}{\sqrt{2}}$, where rms stands for root mean square.
- Thus, $P_{\text{ave}} = I_{\text{rms}} V_{\text{rms}}$.
- Ohm's law for AC is $I_{\text{rms}} = \frac{V_{\text{rms}}}{R}$.
- Expressions for the average power of an AC circuit are $P_{\text{ave}} = I_{\text{rms}} V_{\text{rms}}$, $P_{\text{ave}} = \frac{V_{\text{rms}}^2}{R}$, and $P_{\text{ave}} = I_{\text{rms}}^2 R$, analogous to the expressions for DC circuits.

Conceptual Questions

Exercise:

Problem:

Give an example of a use of AC power other than in the household. Similarly, give an example of a use of DC power other than that supplied by batteries.

Exercise:

Problem:

Why do voltage, current, and power go through zero 120 times per second for 60-Hz AC electricity?

Exercise:

Problem:

You are riding in a train, gazing into the distance through its window. As close objects streak by, you notice that the nearby fluorescent lights make *dashed* streaks. Explain.

Problem Exercises

Exercise:

Problem:

(a) What is the hot resistance of a 25-W light bulb that runs on 120-V AC? (b) If the bulb's operating temperature is $2700^{\circ}\mathrm{C}$, what is its resistance at $2600^{\circ}\mathrm{C}$?

Exercise:

Problem:

Certain heavy industrial equipment uses AC power that has a peak voltage of 679 V. What is the rms voltage?

Solution:

480 V

Exercise:

Problem:

A certain circuit breaker trips when the rms current is 15.0 A. What is the corresponding peak current?

Exercise:

Problem:

Military aircraft use 400-Hz AC power, because it is possible to design lighter-weight equipment at this higher frequency. What is the time for one complete cycle of this power?

Solution:

2.50 ms

Exercise:

Problem:

A North American tourist takes his 25.0-W, 120-V AC razor to Europe, finds a special adapter, and plugs it into 240 V AC. Assuming constant resistance, what power does the razor consume as it is ruined?

Exercise:

Problem:

In this problem, you will verify statements made at the end of the power losses for [link]. (a) What current is needed to transmit 100 MW of power at a voltage of 25.0 kV? (b) Find the power loss in a 1.00 - Ω transmission line. (c) What percent loss does this represent?

Solution:

(a) 4.00 kA

(b) 16.0 MW

(c) 16.0%

Exercise:

Problem:

A small office-building air conditioner operates on 408-V AC and consumes 50.0 kW. (a) What is its effective resistance? (b) What is the cost of running the air conditioner during a hot summer month when it is on 8.00 h per day for 30 days and electricity costs $9.00\ \text{cents}/\text{kW}\cdot\text{h}$?

Exercise:

Problem:

What is the peak power consumption of a 120-V AC microwave oven that draws 10.0 A?

Solution:

2.40 kW

Exercise:

Problem:

What is the peak current through a 500-W room heater that operates on 120-V AC power?

Exercise:

Problem:

Two different electrical devices have the same power consumption, but one is meant to be operated on 120-V AC and the other on 240-V AC. (a) What is the ratio of their resistances? (b) What is the ratio of their currents? (c) Assuming its resistance is unaffected, by what factor will the power increase if a 120-V AC device is connected to 240-V AC?

Solution:

(a) 4.0

(b) 0.50

(c) 4.0

Exercise:

Problem:

Nichrome wire is used in some radiative heaters. (a) Find the resistance needed if the average power output is to be 1.00 kW utilizing 120-V AC. (b) What length of Nichrome wire, having a cross-sectional area of 5.00mm^2, is needed if the operating temperature is $500^{\text{o}}\ \text{C}$? (c) What power will it draw when first switched on?

Exercise:

Problem:

Find the time after $t = 0$ when the instantaneous voltage of 60-Hz AC first reaches the following values: (a) $V_0/2$ (b) V_0 (c) 0.

Solution:

(a) 1.39 ms

(b) 4.17 ms

(c) 8.33 ms

Exercise:

Problem:

(a) At what two times in the first period following $t = 0$ does the instantaneous voltage in 60-Hz AC equal V_{rms}? (b) $-V_{\text{rms}}$?

Glossary

direct current
: (DC) the flow of electric charge in only one direction

alternating current
: (AC) the flow of electric charge that periodically reverses direction

AC voltage
: voltage that fluctuates sinusoidally with time, expressed as $V = V_0 \sin 2\pi ft$, where V is the voltage at time t, V_0 is the peak voltage, and f is the frequency in hertz

AC current
: current that fluctuates sinusoidally with time, expressed as $I = I_0 \sin 2\pi ft$, where I is the current at time t, I_0 is the peak current, and f is the frequency in hertz

rms current
: the root mean square of the current, $I_{\text{rms}} = I_0/\sqrt{2}$, where I_0 is the peak current, in an AC system

rms voltage
: the root mean square of the voltage, $V_{\text{rms}} = V_0/\sqrt{2}$, where V_0 is the peak voltage, in an AC system

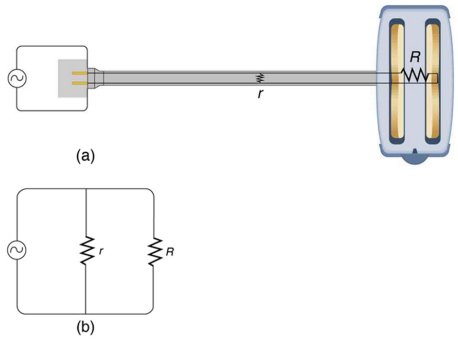
Electric Hazards and the Human Body

- Define thermal hazard, shock hazard, and short circuit.
- Explain what effects various levels of current have on the human body.

There are two known hazards of electricity—thermal and shock. A **thermal hazard** is one where excessive electric power causes undesired thermal effects, such as starting a fire in the wall of a house. A **shock hazard** occurs when electric current passes through a person. Shocks range in severity from painful, but otherwise harmless, to heart-stopping lethality. This section considers these hazards and the various factors affecting them in a quantitative manner. Electrical Safety: Systems and Devices will consider systems and devices for preventing electrical hazards.

Thermal Hazards

Electric power causes undesired heating effects whenever electric energy is converted to thermal energy at a rate faster than it can be safely dissipated. A classic example of this is the **short circuit**, a low-resistance path between terminals of a voltage source. An example of a short circuit is shown in [link]. Insulation on wires leading to an appliance has worn through, allowing the two wires to come into contact. Such an undesired contact with a high voltage is called a *short*. Since the resistance of the short, r, is very small, the power dissipated in the short, $P = V^2/r$, is very large. For example, if V is 120 V and r is $0.100\ \Omega$, then the power is 144 kW, *much* greater than that used by a typical household appliance. Thermal energy delivered at this rate will very quickly raise the temperature of surrounding materials, melting or perhaps igniting them.

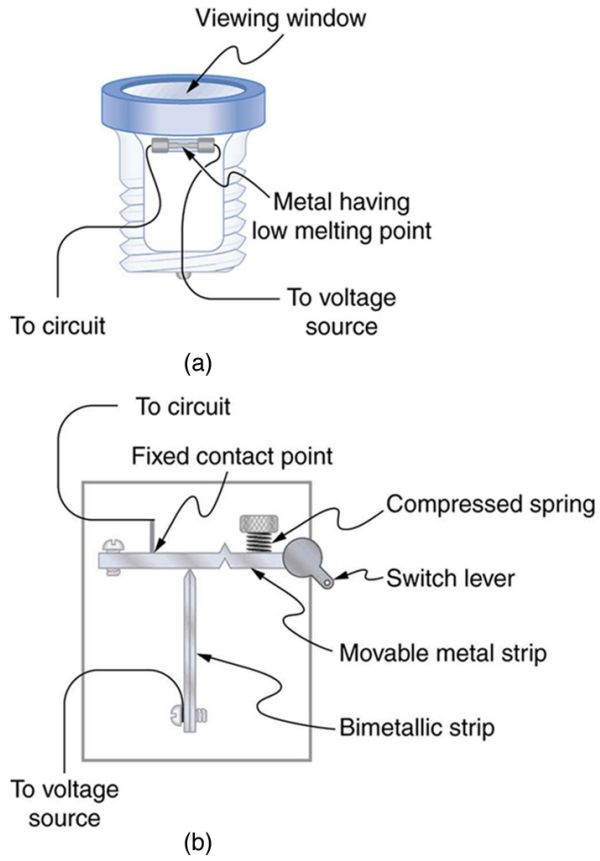


A short circuit is an undesired low-resistance path across a voltage source. (a) Worn insulation on the wires of a toaster allow them to come into contact with a low resistance r. Since $P = V^2/r$, thermal power is created so rapidly that the cord melts or burns. (b) A schematic of the short circuit.

One particularly insidious aspect of a short circuit is that its resistance may actually be decreased due to the increase in temperature. This can happen if the short creates ionization. These charged atoms and molecules are free to move and, thus, lower the resistance r. Since $P = V^2/r$, the power dissipated in the short rises, possibly causing more ionization, more power, and so on. High voltages, such as the 480-V AC used in some industrial applications, lend themselves to this hazard, because higher voltages create higher initial power production in a short.

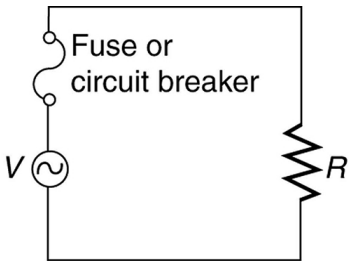
Another serious, but less dramatic, thermal hazard occurs when wires supplying power to a user are overloaded with too great a current. As

discussed in the previous section, the power dissipated in the supply wires is $P = I^2 R_{\text{w}}$, where R_{w} is the resistance of the wires and I the current flowing through them. If either I or R_{w} is too large, the wires overheat. For example, a worn appliance cord (with some of its braided wires broken) may have $R_{\text{w}} = 2.00\ \Omega$ rather than the $0.100\ \Omega$ it should be. If 10.0 A of current passes through the cord, then $P = I^2 R_{\text{w}} = 200$ W is dissipated in the cord—much more than is safe. Similarly, if a wire with a 0.100 - Ω resistance is meant to carry a few amps, but is instead carrying 100 A, it will severely overheat. The power dissipated in the wire will in that case be $P = 1000$ W. Fuses and circuit breakers are used to limit excessive currents. (See [link] and [link].) Each device opens the circuit automatically when a sustained current exceeds safe limits.



(a) A fuse has a metal strip with a low melting point that, when overheated by an excessive

current, permanently breaks the connection of a circuit to a voltage source. (b) A circuit breaker is an automatic but restorable electric switch. The one shown here has a bimetallic strip that bends to the right and into the notch if overheated. The spring then forces the metal strip downward, breaking the electrical connection at the points.



Schematic of a circuit with a fuse or circuit breaker in it. Fuses and circuit breakers act like automatic switches that open when sustained current exceeds desired limits.

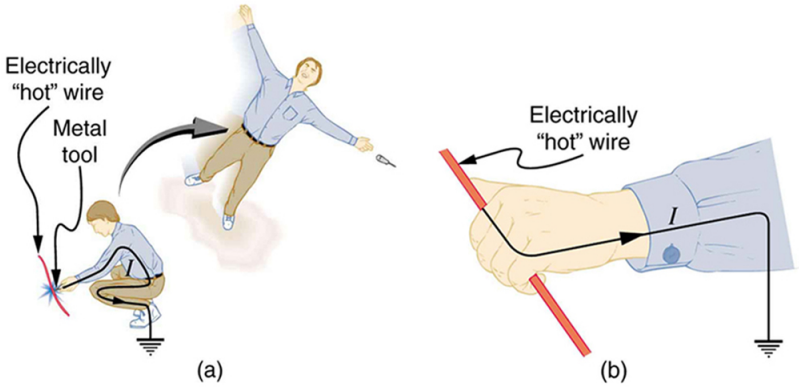
Fuses and circuit breakers for typical household voltages and currents are relatively simple to produce, but those for large voltages and currents experience special problems. For example, when a circuit breaker tries to interrupt the flow of high-voltage electricity, a spark can jump across its points that ionizes the air in the gap and allows the current to continue flowing. Large circuit breakers found in power-distribution systems employ insulating gas and even use jets of gas to blow out such sparks. Here AC is safer than DC, since AC current goes through zero 120 times per second, giving a quick opportunity to extinguish these arcs.

Shock Hazards

Electrical currents through people produce tremendously varied effects. An electrical current can be used to block back pain. The possibility of using electrical current to stimulate muscle action in paralyzed limbs, perhaps allowing paraplegics to walk, is under study. TV dramatizations in which electrical shocks are used to bring a heart attack victim out of ventricular fibrillation (a massively irregular, often fatal, beating of the heart) are more than common. Yet most electrical shock fatalities occur because a current put the heart into fibrillation. A pacemaker uses electrical shocks to stimulate the heart to beat properly. Some fatal shocks do not produce burns, but warts can be safely burned off with electric current (though freezing using liquid nitrogen is now more common). Of course, there are consistent explanations for these disparate effects. The major factors upon which the effects of electrical shock depend are

1. The amount of current I
2. The path taken by the current
3. The duration of the shock
4. The frequency f of the current ($f = 0$ for DC)

[link] gives the effects of electrical shocks as a function of current for a typical accidental shock. The effects are for a shock that passes through the trunk of the body, has a duration of 1 s, and is caused by 60-Hz power.



An electric current can cause muscular contractions with varying effects. (a) The victim is “thrown” backward by involuntary muscle contractions that extend the legs and torso. (b) The victim can’t let go of the wire that is stimulating all the muscles in the hand. Those that close the fingers are stronger than those that open them.

Current (mA)	**Effect**
1	Threshold of sensation
5	Maximum harmless current
10–20	Onset of sustained muscular contraction; cannot let go for duration of shock; contraction of chest muscles may stop breathing during shock

Current (mA)	Effect
50	Onset of pain
100–300+	Ventricular fibrillation possible; often fatal
300	Onset of burns depending on concentration of current
6000 (6 A)	Onset of sustained ventricular contraction and respiratory paralysis; both cease when shock ends; heartbeat may return to normal; used to defibrillate the heart

Effects of Electrical Shock as a Function of Current[footnote]
For an average male shocked through trunk of body for 1 s by 60-Hz AC. Values for females are 60–80% of those listed.

Our bodies are relatively good conductors due to the water in our bodies. Given that larger currents will flow through sections with lower resistance (to be further discussed in the next chapter), electric currents preferentially flow through paths in the human body that have a minimum resistance in a direct path to earth. The earth is a natural electron sink. Wearing insulating shoes, a requirement in many professions, prohibits a pathway for electrons by providing a large resistance in that path. Whenever working with high-power tools (drills), or in risky situations, ensure that you do not provide a pathway for current flow (especially through the heart).

Very small currents pass harmlessly and unfelt through the body. This happens to you regularly without your knowledge. The threshold of sensation is only 1 mA and, although unpleasant, shocks are apparently harmless for currents less than 5 mA. A great number of safety rules take the 5-mA value for the maximum allowed shock. At 10 to 20 mA and above, the current can stimulate sustained muscular contractions much as regular nerve impulses do. People sometimes say they were knocked across the room by a shock, but what really happened was that certain muscles

contracted, propelling them in a manner not of their own choosing. (See [link](a).) More frightening, and potentially more dangerous, is the “can’t let go” effect illustrated in [link](b). The muscles that close the fingers are stronger than those that open them, so the hand closes involuntarily on the wire shocking it. This can prolong the shock indefinitely. It can also be a danger to a person trying to rescue the victim, because the rescuer’s hand may close about the victim’s wrist. Usually the best way to help the victim is to give the fist a hard knock/blow/jar with an insulator or to throw an insulator at the fist. Modern electric fences, used in animal enclosures, are now pulsed on and off to allow people who touch them to get free, rendering them less lethal than in the past.

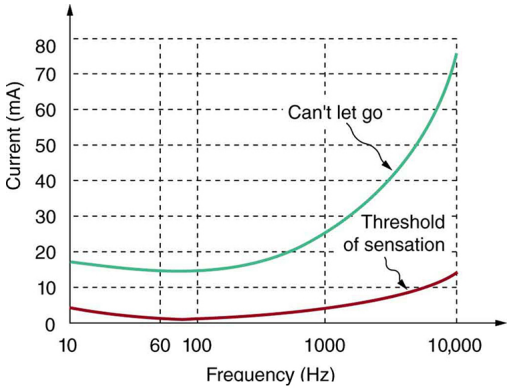
Greater currents may affect the heart. Its electrical patterns can be disrupted, so that it beats irregularly and ineffectively in a condition called “ventricular fibrillation.” This condition often lingers after the shock and is fatal due to a lack of blood circulation. The threshold for ventricular fibrillation is between 100 and 300 mA. At about 300 mA and above, the shock can cause burns, depending on the concentration of current—the more concentrated, the greater the likelihood of burns.

Very large currents cause the heart and diaphragm to contract for the duration of the shock. Both the heart and breathing stop. Interestingly, both often return to normal following the shock. The electrical patterns on the heart are completely erased in a manner that the heart can start afresh with normal beating, as opposed to the permanent disruption caused by smaller currents that can put the heart into ventricular fibrillation. The latter is something like scribbling on a blackboard, whereas the former completely erases it. TV dramatizations of electric shock used to bring a heart attack victim out of ventricular fibrillation also show large paddles. These are used to spread out current passed through the victim to reduce the likelihood of burns.

Current is the major factor determining shock severity (given that other conditions such as path, duration, and frequency are fixed, such as in the table and preceding discussion). A larger voltage is more hazardous, but since $I = \mathrm{V}/\mathrm{R}$, the severity of the shock depends on the combination of voltage and resistance. For example, a person with dry skin has a resistance

of about 200 kΩ. If he comes into contact with 120-V AC, a current $I = (120\text{ V})/(200\text{ k}\Omega) = 0.6\text{ mA}$ passes harmlessly through him. The same person soaking wet may have a resistance of 10.0 kΩ and the same 120 V will produce a current of 12 mA—above the “can’t let go” threshold and potentially dangerous.

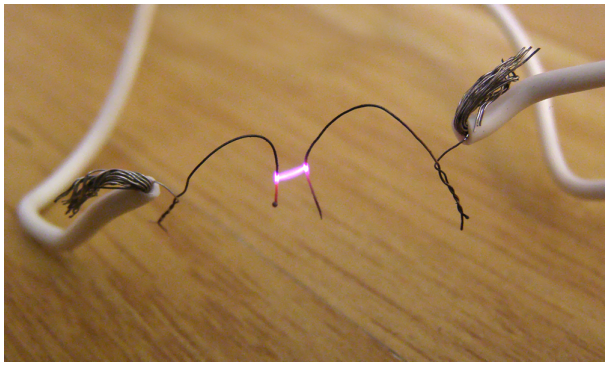
Most of the body’s resistance is in its dry skin. When wet, salts go into ion form, lowering the resistance significantly. The interior of the body has a much lower resistance than dry skin because of all the ionic solutions and fluids it contains. If skin resistance is bypassed, such as by an intravenous infusion, a catheter, or exposed pacemaker leads, a person is rendered **microshock sensitive**. In this condition, currents about 1/1000 those listed in [link] produce similar effects. During open-heart surgery, currents as small as 20 μA can be used to still the heart. Stringent electrical safety requirements in hospitals, particularly in surgery and intensive care, are related to the doubly disadvantaged microshock-sensitive patient. The break in the skin has reduced his resistance, and so the same voltage causes a greater current, and a much smaller current has a greater effect.



Graph of average values for the threshold of sensation and the “can’t let go” current as a function of frequency. The lower the value, the

more sensitive the body is at that frequency.

Factors other than current that affect the severity of a shock are its path, duration, and AC frequency. Path has obvious consequences. For example, the heart is unaffected by an electric shock through the brain, such as may be used to treat manic depression. And it is a general truth that the longer the duration of a shock, the greater its effects. [link] presents a graph that illustrates the effects of frequency on a shock. The curves show the minimum current for two different effects, as a function of frequency. The lower the current needed, the more sensitive the body is at that frequency. Ironically, the body is most sensitive to frequencies near the 50- or 60-Hz frequencies in common use. The body is slightly less sensitive for DC ($f = 0$), mildly confirming Edison's claims that AC presents a greater hazard. At higher and higher frequencies, the body becomes progressively less sensitive to any effects that involve nerves. This is related to the maximum rates at which nerves can fire or be stimulated. At very high frequencies, electrical current travels only on the surface of a person. Thus a wart can be burned off with very high frequency current without causing the heart to stop. (Do not try this at home with 60-Hz AC!) Some of the spectacular demonstrations of electricity, in which high-voltage arcs are passed through the air and over people's bodies, employ high frequencies and low currents. (See [link].) Electrical safety devices and techniques are discussed in detail in Electrical Safety: Systems and Devices.



Is this electric arc dangerous?

The answer depends on the AC frequency and the power involved. (credit: Khimich Alex, Wikimedia Commons)

Section Summary

- The two types of electric hazards are thermal (excessive power) and shock (current through a person).
- Shock severity is determined by current, path, duration, and AC frequency.
- [link] lists shock hazards as a function of current.
- [link] graphs the threshold current for two hazards as a function of frequency.

Conceptual Questions

Exercise:

Problem:

Using an ohmmeter, a student measures the resistance between various points on his body. He finds that the resistance between two points on the same finger is about the same as the resistance between two points on opposite hands—both are several hundred thousand ohms. Furthermore, the resistance decreases when more skin is brought into contact with the probes of the ohmmeter. Finally, there is a dramatic drop in resistance (to a few thousand ohms) when the skin is wet. Explain these observations and their implications regarding skin and internal resistance of the human body.

Exercise:

Problem: What are the two major hazards of electricity?

Exercise:

Problem: Why isn’t a short circuit a shock hazard?

Exercise:

Problem:

What determines the severity of a shock? Can you say that a certain voltage is hazardous without further information?

Exercise:

Problem:

An electrified needle is used to burn off warts, with the circuit being completed by having the patient sit on a large butt plate. Why is this plate large?

Exercise:

Problem:

Some surgery is performed with high-voltage electricity passing from a metal scalpel through the tissue being cut. Considering the nature of electric fields at the surface of conductors, why would you expect most of the current to flow from the sharp edge of the scalpel? Do you think high- or low-frequency AC is used?

Exercise:

Problem:

Some devices often used in bathrooms, such as hairdryers, often have safety messages saying “Do not use when the bathtub or basin is full of water.” Why is this so?

Exercise:

Problem:

We are often advised to not flick electric switches with wet hands, dry your hand first. We are also advised to never throw water on an electric fire. Why is this so?

Exercise:

Problem:

Before working on a power transmission line, linemen will touch the line with the back of the hand as a final check that the voltage is zero. Why the back of the hand?

Exercise:

Problem:

Why is the resistance of wet skin so much smaller than dry, and why do blood and other bodily fluids have low resistances?

Exercise:

Problem:

Could a person on intravenous infusion (an IV) be microshock sensitive?

Exercise:

Problem:

In view of the small currents that cause shock hazards and the larger currents that circuit breakers and fuses interrupt, how do they play a role in preventing shock hazards?

Problem Exercises

Exercise:

Problem:

(a) How much power is dissipated in a short circuit of 240-V AC through a resistance of $0.250\ \Omega$? (b) What current flows?

Solution:

(a) 230 kW

(b) 960 A

Exercise:

Problem:

What voltage is involved in a 1.44-kW short circuit through a 0.100 - Ω resistance?

Exercise:

Problem:

Find the current through a person and identify the likely effect on her if she touches a 120-V AC source: (a) if she is standing on a rubber mat and offers a total resistance of 300 kΩ; (b) if she is standing barefoot on wet grass and has a resistance of only 4000 kΩ.

Solution:

(a) 0.400 mA, no effect

(b) 26.7 mA, muscular contraction for duration of the shock (can't let go)

Exercise:

Problem:

While taking a bath, a person touches the metal case of a radio. The path through the person to the drainpipe and ground has a resistance of 4000 Ω. What is the smallest voltage on the case of the radio that could cause ventricular fibrillation?

Exercise:

Problem:

Foolishly trying to fish a burning piece of bread from a toaster with a metal butter knife, a man comes into contact with 120-V AC. He does not even feel it since, luckily, he is wearing rubber-soled shoes. What is the minimum resistance of the path the current follows through the person?

Solution:

$1.20 \times 10^5\ \Omega$

Exercise:

Problem:

(a) During surgery, a current as small as $20.0\ \mu\text{A}$ applied directly to the heart may cause ventricular fibrillation. If the resistance of the exposed heart is $300\ \Omega$, what is the smallest voltage that poses this danger? (b) Does your answer imply that special electrical safety precautions are needed?

Exercise:

Problem:

(a) What is the resistance of a 220-V AC short circuit that generates a peak power of 96.8 kW? (b) What would the average power be if the voltage was 120 V AC?

Solution:

(a) $1.00\ \Omega$

(b) 14.4 kW

Exercise:

Problem:

A heart defibrillator passes 10.0 A through a patient's torso for 5.00 ms in an attempt to restore normal beating. (a) How much charge passed? (b) What voltage was applied if 500 J of energy was dissipated? (c) What was the path's resistance? (d) Find the temperature increase caused in the 8.00 kg of affected tissue.

Exercise:

Problem: Integrated Concepts

A short circuit in a 120-V appliance cord has a $0.500\text{-}\Omega$ resistance. Calculate the temperature rise of the 2.00 g of surrounding materials, assuming their specific heat capacity is $0.200\ \text{cal/g}\cdot^{\text{o}}\text{C}$ and that it takes 0.0500 s for a circuit breaker to interrupt the current. Is this likely to be damaging?

Solution:

Temperature increases $860^{\text{o}}\ \text{C}$. It is very likely to be damaging.

Exercise:

Problem: Construct Your Own Problem

Consider a person working in an environment where electric currents might pass through her body. Construct a problem in which you calculate the resistance of insulation needed to protect the person from harm. Among the things to be considered are the voltage to which the person might be exposed, likely body resistance (dry, wet, …), and acceptable currents (safe but sensed, safe and unfelt, …).

Glossary

thermal hazard
: a hazard in which electric current causes undesired thermal effects

shock hazard
: when electric current passes through a person

short circuit
: also known as a “short,” a low-resistance path between terminals of a voltage source

microshock sensitive
: a condition in which a person’s skin resistance is bypassed, possibly by a medical procedure, rendering the person vulnerable to electrical shock at currents about 1/1000 the normally required level

DC Circuits Containing Resistors and Capacitors

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain the importance of the time constant τ, and calculate the time constant for a given resistance and capacitance.
- Explain why batteries in a flashlight gradually lose power and the light dims over time.
- Describe what happens to a graph of the voltage across a capacitor over time as it charges.
- Explain how a timing circuit works and list some applications.
- Calculate the necessary speed of a strobe flash needed to "stop" the movement of an object over a particular length.

The information presented in this section supports the following AP® learning objectives and science practices:

- **5.C.3.6** The student is able to determine missing values and direction of electric current in branches of a circuit with both resistors and capacitors from values and directions of current in other branches of the circuit through appropriate selection of nodes and application of the junction rule. **(S.P. 1.4, 2.2)**
- **5.C.3.7** The student is able to determine missing values, direction of electric current, charge of capacitors at steady state, and potential differences within a circuit with resistors and capacitors from values and directions of current in other branches of the circuit. **(S.P. 1.4, 2.2)**

When you use a flash camera, it takes a few seconds to charge the capacitor that powers the flash. The light flash discharges the capacitor in a tiny fraction of a second. Why does charging take longer than discharging? This question and a number of other phenomena that involve charging and discharging capacitors are discussed in this module.

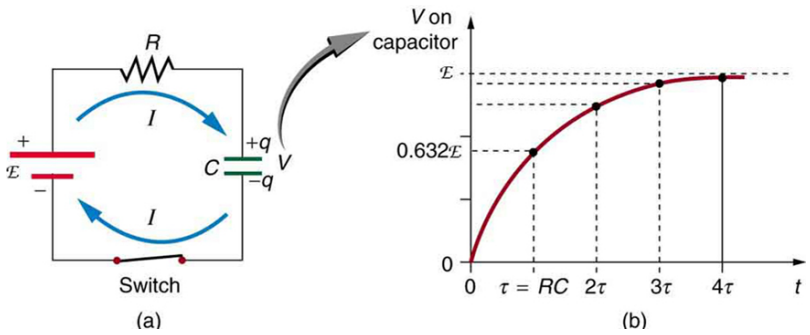
RC Circuits

An **RC circuit** is one containing a **resistor** R and a **capacitor** C. The capacitor is an electrical component that stores electric charge.

[link] shows a simple RC circuit that employs a DC (direct current) voltage source. The capacitor is initially uncharged. As soon as the switch is closed, current flows to and from the initially uncharged capacitor. As charge increases on the capacitor plates, there is increasing opposition to the flow of charge by the repulsion of like charges on each plate.

In terms of voltage, this is because voltage across the capacitor is given by $V_c = Q/C$, where Q is the amount of charge stored on each plate and C is the **capacitance**. This voltage opposes the battery, growing from zero to the maximum emf when fully charged. The current thus decreases from its initial value of $I_0 = \frac{\text{emf}}{R}$ to zero as the voltage on the capacitor reaches the same value as the emf. When there is no current, there is no IR drop, and so the voltage on the capacitor must then equal the emf of the voltage source. This can also be explained with Kirchhoff's second rule (the loop rule), discussed in Kirchhoff's Rules, which says that the algebraic sum of changes in potential around any closed loop must be zero.

The initial current is $I_0 = \frac{\text{emf}}{R}$, because all of the IR drop is in the resistance. Therefore, the smaller the resistance, the faster a given capacitor will be charged. Note that the internal resistance of the voltage source is included in R, as are the resistances of the capacitor and the connecting wires. In the flash camera scenario above, when the batteries powering the camera begin to wear out, their internal resistance rises, reducing the current and lengthening the time it takes to get ready for the next flash.



(a) An RC circuit with an initially uncharged capacitor. Current flows in the direction shown (opposite of electron flow) as soon as the switch is closed. Mutual repulsion of like charges in the capacitor progressively slows the flow as the capacitor is charged, stopping the current when the capacitor is fully charged and $Q = C \cdot \text{emf}$. (b) A graph of voltage across the capacitor versus time, with the switch closing at time $t = 0$. (Note that in the two parts of the figure, the capital script E stands for emf, q stands for the charge stored on the capacitor, and τ is the RC time constant.)

Voltage on the capacitor is initially zero and rises rapidly at first, since the initial current is a maximum. [link](b) shows a graph of capacitor voltage versus time (t) starting when the switch is closed at $t = 0$. The voltage approaches emf asymptotically, since the closer it gets to emf the less current flows. The equation for voltage versus time when charging a capacitor C through a resistor R, derived using calculus, is

Equation:

$$V = \text{emf}(1 - e^{-t/\text{RC}}) \text{ (charging)},$$

where V is the voltage across the capacitor, emf is equal to the emf of the DC voltage source, and the exponential e = 2.718 … is the base of the natural logarithm. Note that the units of RC are seconds. We define
Equation:

$$\tau = \mathrm{RC},$$

where τ (the Greek letter tau) is called the time constant for an RC circuit. As noted before, a small resistance R allows the capacitor to charge faster. This is reasonable, since a larger current flows through a smaller resistance. It is also reasonable that the smaller the capacitor C, the less time needed to charge it. Both factors are contained in $\tau = \mathrm{RC}$.

More quantitatively, consider what happens when $t = \tau = \mathrm{RC}$. Then the voltage on the capacitor is
Equation:

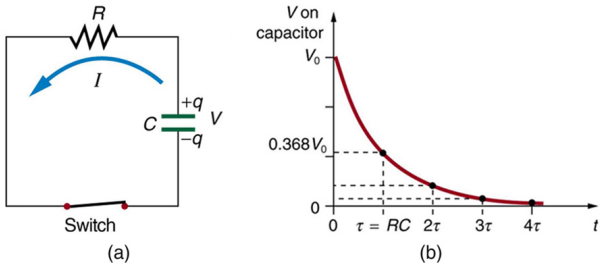
$$V = \mathrm{emf}\left(1 - e^{-1}\right) = \mathrm{emf}(1 - 0.368) = 0.632 \cdot \mathrm{emf}.$$

This means that in the time $\tau = \mathrm{RC}$, the voltage rises to 0.632 of its final value. The voltage will rise 0.632 of the remainder in the next time τ. It is a characteristic of the exponential function that the final value is never reached, but 0.632 of the remainder to that value is achieved in every time, τ. In just a few multiples of the time constant τ, then, the final value is very nearly achieved, as the graph in [link](b) illustrates.

Discharging a Capacitor

Discharging a capacitor through a resistor proceeds in a similar fashion, as [link] illustrates. Initially, the current is $I_0 = \frac{V_0}{R}$, driven by the initial voltage V_0 on the capacitor. As the voltage decreases, the current and hence the rate of discharge decreases, implying another exponential formula for V. Using calculus, the voltage V on a capacitor C being discharged through a resistor R is found to be
Equation:

$$V = V_0\, e^{-t/\mathrm{RC}} (\text{discharging}).$$



(a) Closing the switch discharges the capacitor C through the resistor R. Mutual repulsion of like charges on each plate drives the current. (b) A graph of voltage across the capacitor versus time, with $V = V_0$ at $t = 0$. The voltage decreases exponentially, falling a fixed fraction of the way to zero in each subsequent time constant τ.

The graph in [link](b) is an example of this exponential decay. Again, the time constant is $\tau = \mathrm{RC}$. A small resistance R allows the capacitor to discharge in a small time, since the current is larger. Similarly, a small capacitance requires less time to discharge, since less charge is stored. In the first time interval $\tau = \mathrm{RC}$ after the switch is closed, the voltage falls to 0.368 of its initial value, since $V = V_0 \cdot e^{-1} = 0.368V_0$.

During each successive time τ, the voltage falls to 0.368 of its preceding value. In a few multiples of τ, the voltage becomes very close to zero, as indicated by the graph in [link](b).

Now we can explain why the flash camera in our scenario takes so much longer to charge than discharge; the resistance while charging is significantly greater than while discharging. The internal resistance of the battery accounts for most of the resistance while charging. As the battery ages, the increasing internal resistance makes the charging process even slower. (You may have noticed this.)

The flash discharge is through a low-resistance ionized gas in the flash tube and proceeds very rapidly. Flash photographs, such as in [link], can capture a brief instant of a rapid motion because the flash can be less than a microsecond in duration. Such flashes can be made extremely intense.

During World War II, nighttime reconnaissance photographs were made from the air with a single flash illuminating more than a square kilometer of enemy territory. The brevity of the flash eliminated blurring due to the surveillance aircraft's motion. Today, an important use of intense flash lamps is to pump energy into a laser. The short intense flash can rapidly energize a laser and allow it to reemit the energy in another form.



This stop-motion photograph of a rufous hummingbird (*Selasphorus rufus*) feeding on a flower was obtained with an extremely brief and intense flash of light powered by the discharge of a capacitor through a gas.

(credit: Dean E. Biggins, U.S. Fish and Wildlife Service)

Example:

Integrated Concept Problem: Calculating Capacitor Size—Strobe Lights

High-speed flash photography was pioneered by Doc Edgerton in the 1930s, while he was a professor of electrical engineering at MIT. You might have seen examples of his work in the amazing shots of hummingbirds in motion, a drop of milk splattering on a table, or a bullet penetrating an apple (see [link]). To stop the motion and capture these pictures, one needs a high-intensity, very short pulsed flash, as mentioned earlier in this module.

Suppose one wished to capture the picture of a bullet (moving at $5.0 \times 10^2\ \mathrm{m/s}$) that was passing through an apple. The duration of the flash is related to the RC time constant, τ. What size capacitor would one need in the RC circuit to succeed, if the resistance of the flash tube was $10.0\ \Omega$? Assume the apple is a sphere with a diameter of $8.0 \times 10^{-2}\ \mathrm{m}$.

Strategy

We begin by identifying the physical principles involved. This example deals with the strobe light, as discussed above. [link] shows the circuit for this probe. The characteristic time τ of the strobe is given as $\tau = \mathrm{RC}$.

Solution

We wish to find C, but we don't know τ. We want the flash to be on only while the bullet traverses the apple. So we need to use the kinematic equations that describe the relationship between distance x, velocity v, and time t:

Equation:

$$x = \mathrm{vt} \text{ or } t = \frac{x}{v}.$$

The bullet's velocity is given as 5.0×10^2 m/s, and the distance x is 8.0×10^{-2} m. The traverse time, then, is

Equation:

$$t = \frac{x}{v} = \frac{8.0 \times 10^{-2}\ \text{m}}{5.0 \times 10^2\ \text{m/s}} = 1.6 \times 10^{-4}\ \text{s}.$$

We set this value for the crossing time t equal to τ. Therefore,

Equation:

$$C = \frac{t}{R} = \frac{1.6 \times 10^{-4}\ \text{s}}{10.0\ \Omega} = 16\ \mu\text{F}.$$

(Note: Capacitance C is typically measured in farads, F, defined as Coulombs per volt. From the equation, we see that C can also be stated in units of seconds per ohm.)

Discussion

The flash interval of 160 μs (the traverse time of the bullet) is relatively easy to obtain today. Strobe lights have opened up new worlds from science to entertainment. The information from the picture of the apple and bullet was used in the Warren Commission Report on the assassination of President John F. Kennedy in 1963 to confirm that only one bullet was fired.

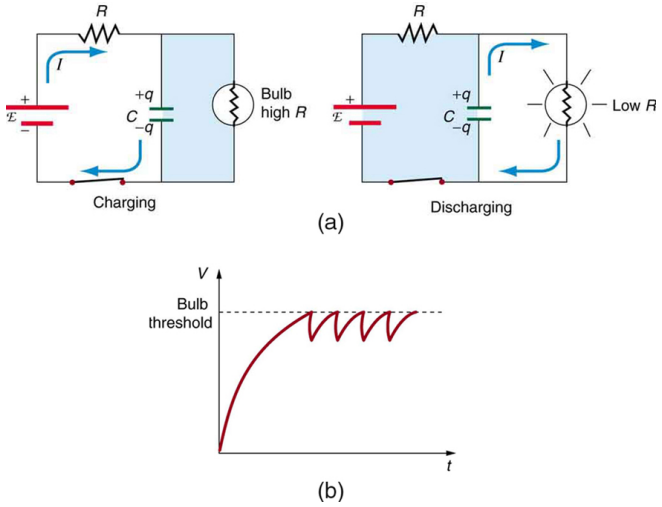
RC Circuits for Timing

RC circuits are commonly used for timing purposes. A mundane example of this is found in the ubiquitous intermittent wiper systems of modern cars. The time between wipes is varied by adjusting the resistance in an RC circuit. Another example of an RC circuit is found in novelty jewelry, Halloween costumes, and various toys that have battery-powered flashing lights. (See [link] for a timing circuit.)

A more crucial use of RC circuits for timing purposes is in the artificial pacemaker, used to control heart rate. The heart rate is normally controlled

by electrical signals generated by the sino-atrial (SA) node, which is on the wall of the right atrium chamber. This causes the muscles to contract and pump blood. Sometimes the heart rhythm is abnormal and the heartbeat is too high or too low.

The artificial pacemaker is inserted near the heart to provide electrical signals to the heart when needed with the appropriate time constant. Pacemakers have sensors that detect body motion and breathing to increase the heart rate during exercise to meet the body's increased needs for blood and oxygen.



(a) The lamp in this RC circuit ordinarily has a very high resistance, so that the battery charges the capacitor as if the lamp were not there. When the voltage reaches a threshold value, a current flows through the lamp that dramatically reduces its resistance, and the capacitor discharges through the lamp as if the battery and charging resistor were not there. Once discharged, the process starts again, with the flash period

determined by the RC constant τ.
(b) A graph of voltage versus time for this circuit.

Example:
Calculating Time: *RC* Circuit in a Heart Defibrillator
A heart defibrillator is used to resuscitate an accident victim by discharging a capacitor through the trunk of her body. A simplified version of the circuit is seen in [link]. (a) What is the time constant if an 8.00-μF capacitor is used and the path resistance through her body is $1.00 \times 10^3\ \Omega$? (b) If the initial voltage is 10.0 kV, how long does it take to decline to 5.00×10^2 V?
Strategy
Since the resistance and capacitance are given, it is straightforward to multiply them to give the time constant asked for in part (a). To find the time for the voltage to decline to 5.00×10^2 V, we repeatedly multiply the initial voltage by 0.368 until a voltage less than or equal to 5.00×10^2 V is obtained. Each multiplication corresponds to a time of τ seconds.
Solution for (a)
The time constant τ is given by the equation $\tau = \mathrm{RC}$. Entering the given values for resistance and capacitance (and remembering that units for a farad can be expressed as s/Ω) gives
Equation:

$$\tau = \mathrm{RC} = (1.00 \times 10^3\ \Omega)(8.00\ \mu F) = 8.00\ \mathrm{ms}.$$

Solution for (b)
In the first 8.00 ms, the voltage (10.0 kV) declines to 0.368 of its initial value. That is:
Equation:

$$V = 0.368 V_0 = 3.680 \times 10^3\ \mathrm{V} \text{ at } t = 8.00\ \mathrm{ms}.$$

(Notice that we carry an extra digit for each intermediate calculation.) After another 8.00 ms, we multiply by 0.368 again, and the voltage is
Equation:

$$\begin{aligned} V\prime &= 0.368V \\ &= (0.368)\left(3.680 \times 10^3 \text{ V}\right) \\ &= 1.354 \times 10^3 \text{ V at } t = 16.0 \text{ ms}. \end{aligned}$$

Similarly, after another 8.00 ms, the voltage is
Equation:

$$\begin{aligned} V\prime\prime &= 0.368V\prime = (0.368)(1.354 \times 10^3 \text{ V}) \\ &= 498 \text{ V at } t = 24.0 \text{ ms}. \end{aligned}$$

Discussion
So after only 24.0 ms, the voltage is down to 498 V, or 4.98% of its original value.Such brief times are useful in heart defibrillation, because the brief but intense current causes a brief but effective contraction of the heart. The actual circuit in a heart defibrillator is slightly more complex than the one in [link], to compensate for magnetic and AC effects that will be covered in Magnetism.

Exercise:
Check Your Understanding

Problem: When is the potential difference across a capacitor an emf?

Solution:

Only when the current being drawn from or put into the capacitor is zero. Capacitors, like batteries, have internal resistance, so their output voltage is not an emf unless current is zero. This is difficult to measure in practice so we refer to a capacitor’s voltage rather than its emf. But the source of potential difference in a capacitor is fundamental and it is an emf.

Note:
PhET Explorations: Circuit Construction Kit (DC only)
An electronics kit in your computer! Build circuits with resistors, light bulbs, batteries, and switches. Take measurements with the realistic ammeter and voltmeter. View the circuit as a schematic diagram, or switch to a life-like view.

Circuit Construction Kit (DC only)

Test Prep for AP Courses

Exercise:

Problem:

A battery is connected to a resistor and an uncharged capacitor. The switch for the circuit is closed at $t = 0$ s.

While the capacitor is being charged, which of the following is
a. true?

a. Current through and voltage across the resistor increase.
b. Current through and voltage across the resistor decrease.
c. Current through and voltage across the resistor first increase and then decrease.

d. Current through and voltage across the resistor first decrease and then increase.

b. When the capacitor is fully charged, which of the following is NOT zero?

a. Current in the resistor.
b. Voltage across the resistor.
c. Current in the capacitor.
d. None of the above.

Exercise:

Problem:

An uncharged capacitor C is connected in series (with a switch) to a resistor R_1 and a voltage source E. Assume E = 24 V, R_1 = 1.2 kΩ and C = 1 mF.

a. What will be the current through the circuit as the switch is closed? Draw a circuit diagram and show the direction of current after the switch is closed. How long will it take for the capacitor to be 99% charged?
b. After full charging, this capacitor is connected in series to another resistor, R_2 = 1 kΩ. What will be the current in the circuit as soon as it's connected? Draw a circuit diagram and show the direction of current. How long will it take for the capacitor voltage to reach 3.24 V?

Solution:

(a) 20 mA, [link], 5.5 s; (b) 24 mA, [link], 2 s

Section Summary

- An RC circuit is one that has both a resistor and a capacitor.
- The time constant τ for an RC circuit is $\tau = \mathrm{RC}$.
- When an initially uncharged ($V_0 = 0$ at $t = 0$) capacitor in series with a resistor is charged by a DC voltage source, the voltage rises, asymptotically approaching the emf of the voltage source; as a function of time,
  Equation:

$$V = \mathrm{emf}(1 - e^{-t/\mathrm{RC}})(\mathrm{charging}).$$

- Within the span of each time constant τ, the voltage rises by 0.632 of the remaining value, approaching the final voltage asymptotically.
- If a capacitor with an initial voltage V_0 is discharged through a resistor starting at $t = 0$, then its voltage decreases exponentially as given by
  Equation:

$$V = V_0 e^{-t/\mathrm{RC}}(\mathrm{discharging}).$$

- In each time constant τ, the voltage falls by 0.368 of its remaining initial value, approaching zero asymptotically.

Conceptual questions

Exercise:

Problem:

Regarding the units involved in the relationship $\tau = \mathrm{RC}$, verify that the units of resistance times capacitance are time, that is, $\Omega \cdot \mathrm{F} = \mathrm{s}$.

Exercise:

Problem:

The RC time constant in heart defibrillation is crucial to limiting the time the current flows. If the capacitance in the defibrillation unit is fixed, how would you manipulate resistance in the circuit to adjust the RC constant τ? Would an adjustment of the applied voltage also be needed to ensure that the current delivered has an appropriate value?

Exercise:

Problem:

When making an ECG measurement, it is important to measure voltage variations over small time intervals. The time is limited by the RC constant of the circuit—it is not possible to measure time variations shorter than RC. How would you manipulate R and C in the circuit to allow the necessary measurements?

Exercise:

Problem:

Draw two graphs of charge versus time on a capacitor. Draw one for charging an initially uncharged capacitor in series with a resistor, as in the circuit in [link], starting from $\mathrm{t} = 0$. Draw the other for discharging a capacitor through a resistor, as in the circuit in [link], starting at $\mathrm{t} = 0$, with an initial charge Q_0. Show at least two intervals of τ.

Exercise:

Problem:

When charging a capacitor, as discussed in conjunction with [link], how long does it take for the voltage on the capacitor to reach emf? Is this a problem?

Exercise:

Problem:

When discharging a capacitor, as discussed in conjunction with [link], how long does it take for the voltage on the capacitor to reach zero? Is this a problem?

Exercise:

Problem:

Referring to [link], draw a graph of potential difference across the resistor versus time, showing at least two intervals of τ. Also draw a graph of current versus time for this situation.

Exercise:

Problem:

A long, inexpensive extension cord is connected from inside the house to a refrigerator outside. The refrigerator doesn't run as it should. What might be the problem?

Exercise:

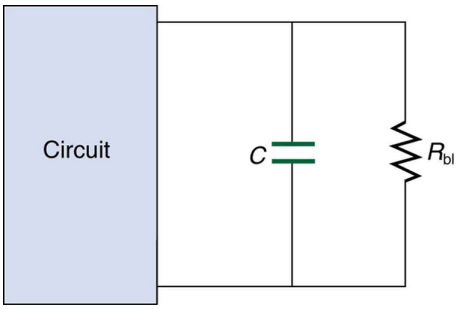
Problem:

In [link], does the graph indicate the time constant is shorter for discharging than for charging? Would you expect ionized gas to have low resistance? How would you adjust R to get a longer time between flashes? Would adjusting R affect the discharge time?

Exercise:

Problem:

An electronic apparatus may have large capacitors at high voltage in the power supply section, presenting a shock hazard even when the apparatus is switched off. A "bleeder resistor" is therefore placed across such a capacitor, as shown schematically in [link], to bleed the charge from it after the apparatus is off. Why must the bleeder resistance be much greater than the effective resistance of the rest of the circuit? How does this affect the time constant for discharging the capacitor?



A bleeder resistor R_{bl} discharges the capacitor in this electronic device once it is switched off.

Problem Exercises

Exercise:

Problem:

The timing device in an automobile's intermittent wiper system is based on an RC time constant and utilizes a $0.500\text{-}\mu\text{F}$ capacitor and a variable resistor. Over what range must R be made to vary to achieve time constants from 2.00 to 15.0 s?

Solution:

range 4.00 to 30.0 $\text{M}\Omega$

Exercise:

Problem:

A heart pacemaker fires 72 times a minute, each time a 25.0-nF capacitor is charged (by a battery in series with a resistor) to 0.632 of its full voltage. What is the value of the resistance?

Exercise:

Problem:

The duration of a photographic flash is related to an RC time constant, which is $0.100\ \mu\mathrm{s}$ for a certain camera. (a) If the resistance of the flash lamp is $0.0400\ \Omega$ during discharge, what is the size of the capacitor supplying its energy? (b) What is the time constant for charging the capacitor, if the charging resistance is $800\ \mathrm{k}\Omega$?

Solution:

(a) $2.50\ \mu\mathrm{F}$

(b) 2.00 s

Exercise:

Problem:

A 2.00- and a $7.50\text{-}\mu\mathrm{F}$ capacitor can be connected in series or parallel, as can a 25.0- and a $100\text{-}\mathrm{k}\Omega$ resistor. Calculate the four RC time constants possible from connecting the resulting capacitance and resistance in series.

Exercise:

Problem:

After two time constants, what percentage of the final voltage, emf, is on an initially uncharged capacitor C, charged through a resistance R?

Solution:

86.5%

Exercise:

Problem:

A $500\text{-}\Omega$ resistor, an uncharged $1.50\text{-}\mu\text{F}$ capacitor, and a 6.16-V emf are connected in series. (a) What is the initial current? (b) What is the RC time constant? (c) What is the current after one time constant? (d) What is the voltage on the capacitor after one time constant?

Exercise:

Problem:

A heart defibrillator being used on a patient has an RC time constant of 10.0 ms due to the resistance of the patient and the capacitance of the defibrillator. (a) If the defibrillator has an $8.00\text{-}\mu\text{F}$ capacitance, what is the resistance of the path through the patient? (You may neglect the capacitance of the patient and the resistance of the defibrillator.) (b) If the initial voltage is 12.0 kV, how long does it take to decline to 6.00×10^2 V?

Solution:

(a) $1.25\ \text{k}\Omega$

(b) 30.0 ms

Exercise:

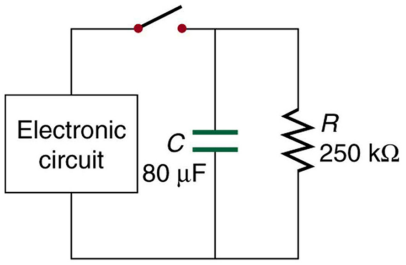
Problem:

An ECG monitor must have an RC time constant less than 1.00×10^2 μs to be able to measure variations in voltage over small time intervals. (a) If the resistance of the circuit (due mostly to that of the patient's chest) is $1.00\ \text{k}\Omega$, what is the maximum capacitance of the circuit? (b) Would it be difficult in practice to limit the capacitance to less than the value found in (a)?

Exercise:

Problem:

[link] shows how a bleeder resistor is used to discharge a capacitor after an electronic device is shut off, allowing a person to work on the electronics with less risk of shock. (a) What is the time constant? (b) How long will it take to reduce the voltage on the capacitor to 0.250% (5% of 5%) of its full value once discharge begins? (c) If the capacitor is charged to a voltage V_0 through a 100-Ω resistance, calculate the time it takes to rise to $0.865V_0$ (This is about two time constants.)



Solution:

(a) 20.0 s

(b) 120 s

(c) 16.0 ms

Exercise:

Problem:

Using the exact exponential treatment, find how much time is required to discharge a 250-μF capacitor through a 500-Ω resistor down to 1.00% of its original voltage.

Exercise:

Problem:

Using the exact exponential treatment, find how much time is required to charge an initially uncharged 100-pF capacitor through a 75.0-MΩ resistor to 90.0% of its final voltage.

Solution:

$1.73 \times 10^{-2}\ \text{s}$

Exercise:

Problem: **Integrated Concepts**

If you wish to take a picture of a bullet traveling at 500 m/s, then a very brief flash of light produced by an RC discharge through a flash tube can limit blurring. Assuming 1.00 mm of motion during one RC constant is acceptable, and given that the flash is driven by a $600\text{-}\mu\text{F}$ capacitor, what is the resistance in the flash tube?

Solution:

$3.33 \times 10^{-3}\ \Omega$

Exercise:

Problem: **Integrated Concepts**

A flashing lamp in a Christmas earring is based on an RC discharge of a capacitor through its resistance. The effective duration of the flash is 0.250 s, during which it produces an average 0.500 W from an average 3.00 V. (a) What energy does it dissipate? (b) How much charge moves through the lamp? (c) Find the capacitance. (d) What is the resistance of the lamp?

Exercise:

Problem: **Integrated Concepts**

A 160-μF capacitor charged to 450 V is discharged through a $31.2\text{-k}\Omega$ resistor. (a) Find the time constant. (b) Calculate the temperature increase of the resistor, given that its mass is 2.50 g and its specific heat is $1.67\ \frac{\text{kJ}}{\text{kg} \cdot {}^{\circ}\text{C}}$, noting that most of the thermal energy is retained in

the short time of the discharge. (c) Calculate the new resistance, assuming it is pure carbon. (d) Does this change in resistance seem significant?

Solution:

(a) 4.99 s

(b) $3.87^{\circ}\mathrm{C}$

(c) $31.1\ \mathrm{k}\Omega$

(d) No

Exercise:

Problem: Unreasonable Results

(a) Calculate the capacitance needed to get an RC time constant of 1.00×10^{3} s with a $0.100\text{-}\Omega$ resistor. (b) What is unreasonable about this result? (c) Which assumptions are responsible?

Exercise:

Problem: Construct Your Own Problem

Consider a camera’s flash unit. Construct a problem in which you calculate the size of the capacitor that stores energy for the flash lamp. Among the things to be considered are the voltage applied to the capacitor, the energy needed in the flash and the associated charge needed on the capacitor, the resistance of the flash lamp during discharge, and the desired RC time constant.

Exercise:

Problem: Construct Your Own Problem

Consider a rechargeable lithium cell that is to be used to power a camcorder. Construct a problem in which you calculate the internal

resistance of the cell during normal operation. Also, calculate the minimum voltage output of a battery charger to be used to recharge your lithium cell. Among the things to be considered are the emf and useful terminal voltage of a lithium cell and the current it should be able to supply to a camcorder.

Glossary

RC circuit
: a circuit that contains both a resistor and a capacitor

capacitor
: an electrical component used to store energy by separating electric charge on two opposing plates

capacitance
: the maximum amount of electric potential energy that can be stored (or separated) for a given electric potential